

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comugen Ltd.

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

cleic - nucleic search, using sw model

December 9, 2003, 01:21:55 : Search time 2842 Seconds

(without alignments)  
9817.159 Million cell updates/sec

US-10-035-914-1

1 aaatccgccttccttcctt.....gaacagctccgtgactgag 682

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

2888711 seqs, 2045481386 residues

number of hits satisfying chosen parameters: 5777422

DB seq length: 0

DB seq length: 2000000000

Maximum Match 0%

Listing first 45 summaries

GenBml:\*

1: gb\_ba:\*

2: gb\_hcg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_da:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_un:\*

29: em\_vl:\*

30: em\_hcg\_hum:\*

31: em\_hcg\_inv:\*

32: em\_hcg\_other:\*

33: em\_hcg\_mus:\*

34: em\_hcg\_pln:\*

35: em\_hcg\_rpd:\*

36: em\_hcg\_vit:\*

37: em\_hcg\_vit:\*

38: em\_hcg\_hum:\*

39: em\_hcg\_mus:\*

40: em\_hcg\_pln:\*

41: em\_hcg\_rpd:\*

42: em\_hcg\_vit:\*

43: em\_hcg\_vit:\*

44: em\_hcg\_vit:\*

45: em\_hcg\_vit:\*

Result No. Score Match Length DB ID

Description

1	682	100.0	1234	10	MMMD3MRN	X59047 M. musculus
2	682	100.0	1510	10	BC011433	BC011433 Mus muscu
3	614.8	90.1	1303	6	AX700189	AX700189 Sequence
4	614.8	90.1	1303	10	RNU19894	U19894 Rattus norv
5	535.4	78.5	711	9	AF116599	AF116599 Chloroceb
6	534.8	78.4	1333	9	BC002978	BC002978 Homo sapi
7	534.8	78.4	1496	6	AX337094	AX337094 Sequence
8	534.8	78.4	1496	6	AX658153	AX658153 Sequence
9	534.8	78.4	1496	9	HUMTAPAL	M13680 Human 26-kD
10	530.6	77.8	711	9	SOE250197	AJ250197 Saguinus
11	527.4	77.3	711	9	AF116600	AF116600 Pan trogl
12	484.2	71.0	633	9	AF274885	AF274885 Saguinus
13	290.4	42.6	1329	5	BC041217	BC041217 Xenopus 1
14	277.4	40.7	1300	5	AF295377	AF295377 Dario rer
15	270	39.6	436	6	AR240547	AR240547 Sequence
16	219.2	32.1	429	6	AX198983	AX198983 Sequence
17	217.8	31.9	316	6	AX209509	AX209509 Sequence
18	213.4	31.3	311	6	AR272638	AR272638 Sequence
19	213.4	31.3	311	6	AR276219	AR276219 Sequence
20	213.4	31.3	311	6	AX062740	AX062740 Sequence
21	213.4	31.3	311	6	AX367657	AX367657 Sequence
22	195.2	28.6	357	6	AX004409	AX004409 Sequence
23	195.2	28.6	357	6	BD077115	BD077115 Hepatitis
24	190.8	28.0	348	6	AX004411	AX004411 Sequence
25	190.8	28.0	348	6	BD077116	BD077116 Hepatitis
26	184	27.0	1513	5	AY152677	AY152677 Petromyzo
27	165	24.2	1240	4	BOVCD9ANT	M61720 Bos taurus
28	155.4	22.8	1227	4	AY072785	AY072785 Sus scrof
29	153.8	22.6	1024	4	CATPCD9	D30786 Felis catus
30	153.8	22.6	1150	4	CATCDAN	L35275 Felis catus
31	153	22.4	1054	4	AF525029	AF525029 Sus scrof
32	152.6	22.4	687	6	EO5732	EO5732 cDNA encodi
33	151	22.1	687	6	AR016440	AR016440 Sequence
34	151	22.1	687	6	113743	113743 Sequence 2
35	151	22.1	687	6	HUMP24CD9A	L34068 Human antilg
36	151	22.1	1120	6	AR016441	AR016441 Sequence
37	151	22.1	1120	6	BD176845	BD176845 A method
38	151	22.1	1120	9	AGMDRAP27	D10726 African gre
39	151	22.1	1120	6	HSMRP1	X60111 H. sapiens m
40	151	22.1	1192	6	AX409704	AX409704 Sequence
41	151	22.1	1192	6	AX658227	AX658227 Sequence
42	151	22.1	1192	9	HUMANTCD9	M38690 Human CD9 a
43	151	22.1	66877	2	AC116062	AC116062 Rattus no
44	149.4	21.9	1120	6	113744	113744 Sequence 3
45	149.4	21.9	1166	9	BC011988	BC011988 Homo sapi

ALIGNMENTS

RESULT 1  
LOCUS MMMD3MRN 1234 bp mRNA linear ROD 27-OCT-1992  
DEFINITION M. musculus MD3 mRNA.  
ACCESSION X59047  
VERSION X59047.1 GI:53037  
KEYWORDS cell surface antigen.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1234)  
AUTHORS Duff, K. and Parsons, J.  
TITLE Mouse Homologue of human TAPA-1 mRNA  
JOURNAL Unpublished



gene /lab host="DH10B"  
/note="Vector: pCMV-SPORT6"  
1. 1510  
/gene="Cd81"  
/note="synonym: CD81"  
/db\_xref="LocusID:12520"  
/db\_xref="MGI:1096398"  
207. 917  
/codon\_start=1  
/product="CD 81 antigen"  
/protein\_id="AAH1433.1"  
/db\_xref="GI:15030313"  
/db\_xref="LocusID:12520"  
/translation="MGVEGCTKCIKYLFFVNFVFLAGVILGVALMLRHPDPTSL  
LYLEAGNKPAPNPFYVGIYILAVGAVMFMFPGICVGLIOESQCLGPTFLVILF  
ACEVAAGIWPVNDQIADKVKOPYDQALQAVMDNNAKAVKTHETLWCCSN  
ALTLTUTTIIRNSLCRSGNITLFLQDQCHQKIDELPSKLYLIGIALVAVIMIF  
EMLSWLCGGINSSVY"

CDS  
COUNT 340 a 400 c 360 g 410 t  
N

Match 100.0%; Score 682; DB 10; Length 1510;  
Local Similarity 100.0%; Pred. No. 9.2e-182;  
ches 682; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 AAATACCTGCTCTGCTCTCAATTTGTCCTGCTGCTGAGGCGTATCTAGT 60  
237 AAATACCTGCTCTGCTCTCAATTTGTCCTGCTGCTGAGGCGTATCTAGT 296  
61 GTAGCTCTGCTGCTGCTCAATTCACAGACCAAGCTGCTGATCTGGAAGT 120  
297 GTAGCTCTGCTGCTGCTCAATTCACAGACCAAGCTGCTGATCTGGAAGT 356  
121 AACAAACCGGACCAACACCTCTAGTGGGATCTACATTTCTCATGCTGGAGCT 180  
357 AACAAACCGGACCAACACCTCTAGTGGGATCTACATTTCTCATGCTGGAGCT 416  
181 GTAGATGTTTGTAGGCTCTGAGGCTGATGAGGCGATCCAGAGTCCAGTCTG 240  
417 GTAGATGTTTGTAGGCTCTGAGGCTGATGAGGCGATCCAGAGTCCAGTCTG 476  
241 CTGGGAGCTTTCTTCACTGCTGCTGATCTGTTTCCCTGAGGCTGCTGAGGCT 300  
477 CTGGGAGCTTTCTTCACTGCTGCTGATCTGTTTCCCTGAGGCTGCTGAGGCT 536  
301 TGGGGCTTGTAAACAAAGACGATGCGCAAGATGTGAAGCATTTATGACAGGCC 360  
537 TGGGGCTTGTAAACAAAGACGATGCGCAAGATGTGAAGCATTTATGACAGGCC 596  
361 CTTCAGCAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
597 CTTCAGCAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 656  
421 CATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
657 CATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 716  
481 AGGAACAGCTGTGCTGCTGAGGCGCAACATCTACACCCCTTACTGACAGAAATGT 540  
717 AGGAACAGCTGTGCTGCTGAGGCGCAACATCTACACCCCTTACTGACAGAAATGT 776  
541 CATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
777 CATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 836  
601 GTGAGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660  
837 GTGAGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 896  
661 CGGAACAGCTCGGTATGAG 682  
897 CGGAACAGCTCGGTATGAG 918

RESULT 3  
AX700189 1303 bp DNA linear PAT 03-Apr-2003  
Sequence 75 from Patent EP1264298.  
AX700189  
AX700189.1 GI:29536100  
KEYWORDS  
SOURCE  
ORGANISM  
Rattus norvegicus (Norway rat)  
Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
REFERENCE  
AUTHORS  
TITILE  
JOURNAL  
WARNER-LAMBERT COMPANY (US)  
Location/Qualifiers  
1. 1303  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/note="Cd81 antigen"

BASE COUNT 257 a 375 c 347 g 324 t  
ORIGIN

Query Match 90.1%; Score 614.8; DB 6; Length 1303;  
Best Local Similarity 93.8%; Pred. No. 9.6e-163;  
Matches 640; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

1 AAATACCTGCTCTGCTCTCAATTTGTCCTGCTGCTGAGGCGTATCTAGT 60  
253 AAATACCTGCTCTGCTCTCAATTTGTCCTGCTGCTGAGGCGTATCTAGT 312  
61 GTAGCTCTGCTGCTGCTCAATTCACAGACCAAGCTGCTGATCTGGAAGT 120  
313 GTAGCTCTGCTGCTGCTCAATTCACAGACCAAGCTGCTGATCTGGAAGT 372  
121 AACAAACCGGACCAACACCTCTAGTGGGATCTACATTTCTCATGCTGGAGCT 180  
373 GACAAACGACACCTAGACCTTCTATGAGGATCTACATTTCTCATGCTGGAGCT 432  
181 GTAGATGTTTGTAGGCTCTGAGGCTGATGAGGCGATCCAGAGTCCAGTCTG 240  
433 GTAGATGTTTGTAGGCTCTGAGGCTGATGAGGCGATCCAGAGTCCAGTCTG 492  
241 CTGGGAGCTTTCTTCACTGCTGCTGATCTGTTTCCCTGAGGCTGCTGAGGCT 300  
493 CTGGGAGCTTTCTTCACTGCTGCTGATCTGTTTCCCTGAGGCTGCTGAGGCT 552  
301 TGGGGCTTGTAAACAAAGACGATGCGCAAGATGTGAAGCATTTATGACAGGCC 360  
553 TGGGGCTTGTAAACAAAGACGATGCGCAAGATGTGAAGCATTTATGACAGGCC 612  
493 CTTCAGCAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 672  
361 CTTCAGCAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 732  
421 CATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
613 CATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 672  
421 CATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
673 CATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 732  
481 AGGAACAGCTGTGCTGCTGAGGCGCAACATCTACACCCCTTACTGACAGAAATGT 540  
733 AGGAACAGCTGTGCTGCTGAGGCGCAACATCTACACCCCTTACTGACAGAAATGT 792  
541 CATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
793 CATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 852  
601 GTGAGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660

853 GTGTAGCTGTCTATATATCTTCAGATGATTTGACGATGCTGTCTGTGGCATC 912  
 661 CGGAACAGCTCCGTGTACTGAG 682  
 913 CGGAACAGCTCCGTGTACTGAG 934

1303 bp mRNA linear ROD 25-MAR-1997  
 Rattus norvegicus target of the antiproliferative antibody mRNA,  
 complete cds.  
 U19894  
 U19894.1 GI:1142641

Rattus norvegicus (Norway rat)  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.

1 (bases 1 to 1303)  
 Geisert, E. B. Jr., Murphy, T. P., Irwin, M. H. and Larjava, H.  
 A novel cell adhesion molecule, G-CAM, found on cultured rat glia  
 Neurosci. Lett. 133 (2), 262-266 (1991)

2 (bases 1 to 1303)  
 Irwin, M. H. and Geisert, E. B. Jr.  
 The upregulation of a glial cell surface antigen at the astrocytic  
 scar in the rat  
 Neurosci. Lett. 154 (1-2), 57-60 (1993)

3 (bases 1 to 1303)  
 Geisert, E. B. Jr., Yang, L. and Irwin, M. H.  
 Astrocyte growth, reactivity, and the target of the  
 antiproliferative antibody, TAPA  
 J. Neurosci. 16 (17), 5478-5487 (1996)

4 (bases 1 to 1303)  
 Geisert, E. B.  
 Direct Submission  
 Submitted (17-JAN-1995) Eldon B. Geisert Jr., Department of Anatomy  
 and Neurobiology, University of Tennessee, Memphis, 855 Monroe  
 Ave., Memphis, TN 38163, USA  
 Location/Qualifiers  
 1. 1303  
 /organism="Rattus norvegicus"  
 /mol\_type="mRNA"  
 /strain="Sprague-Dawley"  
 /db\_xref="taxon:10116"  
 /clone="1.1, sp1, sp2"  
 /sex="Male"  
 /cell\_type="astrocyte (Type 1 and Type 2)"  
 /tissue\_type="cerebral cortex"  
 /clone\_lib="Astrocyte cDNA lambda gt11 (Clontech No.  
 RL1042B)"  
 1..222  
 223..933  
 /note="TAPA"  
 /codon\_start=1  
 /product="target of the antiproliferative antibody"  
 /protein\_id="AAC53103.1"  
 /db\_xref="GI:1142642"  
 /translation="MGVRCCKKIKYLLPVNPPVPLAGVILGVALLRDPOTTL  
 LYLRLGDPASSTPTVGIYIILANGAVMMVGFAGCTGAIQESQCLGFTCLVILP  
 ACEVAAGTGWGVNDQAKDVKQPTDQALQAVMDNDANNAKAVVKTHTLNCGSN  
 TLTLTTLVNLNSLCPSSSSNSTQLKEDCHKIDELFSGLYILGIAIVAVAMIP  
 EMILSMVLCGGRINSYV"  
 934..1303  
 257 a 375 c 347 g 324 t

Query Match 90.1%; Score 614.8; DB 10; Length 1303;  
 Best Local Similarity 93.8%; Pred. No. 9,66-163;  
 Matches 640; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

1 AAATACCTGCTCTTGTCTTCAATTTGCTTCTTGTGCTGTGAGCGGTATCTTAGT 60  
 253 AAATACCTGCTCTTGTCTTCAATTTGCTTCTTGTGCTGTGAGCGGTATCTTAGT 312  
 61 GTAGCTGTGTGCTGCGCATGATCCAGACAGAGCTGTGTACTGTGAATCTGGA 120  
 313 GTAGCTGTGTGCTGCGCATGATCCAGACAGAGCTGTGTACTGTGTGAATCTGGA 372  
 121 AACAAACCGGACCCCAACACCTTCTAGCTGCGGATCTACATTTCTCATTTGCTG 180  
 373 GACAAACAGACACCTTACACCTTCTATGTGCGCATCTACATTTCTCATTTGCTG 432  
 181 GTATGATGTTTGTAGGCTTCTGCGGCTGTATGCGGCGCATCCAGAGTCTGCTG 240  
 433 GTATGATGTTTGTAGGCTTCTGCGGCTGTATGCGGCGCATCCAGAGTCTGCTG 492  
 241 CTGCGGACGTTCTTCACTGCTGTGTATCTGTTGCTGTGTAGGTGTGCGGATC 300  
 493 CTGCGGACGTTCTTCACTGCTGTGTATCTGTTGCTGTGTAGGTGTGCGGATC 552  
 301 TGGGCGCTTCTTAAACAAAGACAGATCGCCAGAGATGAGAGAGTTCTATGACAGG 360  
 553 TGGGCGCTTCTTAAACAAAGACAGATCGCCAGAGATGAGAGAGTTCTATGACAGG 612  
 361 CTTCAGCAAGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
 613 CTTCAGCAAGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 672  
 421 CATGAGAGCTTCACTGTTGTGCTTCAAGCACTGACAGCACTGATCACTACATCTG 480  
 673 CATGAGAGCTTCACTGTTGTGCTTCAAGCACTGACAGCACTGATCACTACATCTG 732  
 481 AGGAACAGCTGTGCTGCTGCGGAGCACTACATCACTGACAGCACTGATGATG 540  
 733 AGGAACAGCTGTGCTGCTGCGGAGCACTACATCACTGACAGCACTGATGATG 792  
 541 CATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
 793 CATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 852  
 601 GTGTAGCTGTCTATATATCTTCAGATGATTTGACGATGCTGTGTGGCATC 660  
 853 GTGTAGCTGTCTATATATCTTCAGATGATTTGACGATGCTGTGTGGCATC 912

RESULT 5  
 AF116599 711 bp mRNA linear PRI 01-FEB-1999  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

Cercopithecus aethiops (African green monkey)  
 Cercopithecus aethiops  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 Cercopithecinae; Cercopithecus.

REFERENCE  
 1 (bases 1 to 711)  
 Levy, S. and Kuo, C. C.  
 African green monkey CD81 cDNA  
 Unpublished  
 2 (bases 1 to 711)  
 Levy, S. and Kuo, C. C.  
 Direct Submission



[illegible]



Match	Local Similarity	77.8%	Score 530.6	DB 9	Length 711
Local Similarity	86.2%	Pred. No. 6.2e-139			
Mismatches	587	Conservative	0	Mismatches	94
				Indels	0
				Gaps	0
1	AAATACCTGCTCTTCTGCTCAATTCGTTCTTCTGAGCTGAGGCGTGAATCTAGGT	60			
31	AAGTACCTGCTCTTCTGCTCTTCACTTCGTTCTTCTGAGCTGAGGCGTGAATCTAGGT	90			
61	GTAAGCTCTGAGTGGTGCATGATATCCAGACCAACGAGCTGCTATCTAGGAATCGGA	120			
91	GAGGCTCTGAGTGGTCCGCAATGAGCCCGAAGACCAACCACTCTGTTATCTGAGAGCTGGGA	150			

```

FEATURES
    source
        location/Qualifiers
            1..711
                /organism="Pan troglodytes"
                /mol_type="mRNA"
                /db_xref="taxon:9598"
                /cell_line="Janice"
                /note="EBV-transformed cells; amplified using Homo sapiens
                CD81-derived primers"
            1..711
                /codon_start=1
                /product="CD81"
                /protein_id="AA011440.1"
    CDS

```











claim 3, Page 214; 274pp; German.

his invention describes novel nucleic acid sequences that are highly expressed in normal ovary tissue. Artificial chromosomes and cosmid clones containing the sequences can be used as gene transfer vehicles. The sequences can be used to produce DNA fragments containing full-length genes. Host cells transfected with the sequences can be used to produce polypeptides or polypeptide fragments, which can be used to screen phase displays for polypeptides that bind to them, or as tools for identifying agents active against ovarian cancer, or to prepare reagents for treating ovarian cancer. The cDNA sequences can be used to obtain genomic clones, their promoters, enhancers, silencers, exon structures, intron structures and their splice variants. AA241222-241224 represent cDNA sequences derived from normal human ovarian tissue and which encode the protein fragments represented in AA159724-Y59837.

sequence 1341 BP; 241 A; 415 C; 375 G; 310 T; 0 other;

Y Match 78.4%; Score 534.8; DB 20; Length 1341;  
Local Similarity 86.5%; Pred. 1.4e-150;  
hes 590; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

```

1  AATATCTGCTCTTCTTCAATTCCTTCTGCTGAGGCTGATCTAGT 60
103 AAGTACCTGCTCTTCTTCAATTCCTTCTGCTGAGGCTGATCTAGT 162
61  GTAGCTGTGCTGCTGATGATGATGATGATGATGATGATGATGAT 120
163 GTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 222
121 AACAAACCGGACCCCAACCTTCTACCTGAGGATCTACATTCATGCTGAGGCT 180
223 GACAAACCGGACCCCAACCTTCTACCTGAGGATCTACATTCATGCTGAGGCT 282
181 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
283 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 342
241 GTGAGGAGCTTCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
343 GTGAGGAGCTTCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 402
301 TGGGGCTTCTTAAACAAAGACAGATGCTGCTGCTGCTGCTGCTGCTGCT 360
403 TGGGGCTTCTTAAACAAAGACAGATGCTGCTGCTGCTGCTGCTGCTGCT 462
361 CTTCAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
463 CTTCAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 522
421 CATGAGAGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
523 CATGAGAGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 582
481 AGGAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
583 AAGAAACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 642
541 CATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
643 CACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 702
601 GTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
703 GTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 762
661 CGGAACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 682
763 CGGAACAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 784

```

ID	ABL69266 standard; DNA; 1496 BP.
XX	
AC	ABL69266;
XX	
DT	15-MAY-2002 (first entry)
XX	
DE	Prostate cancer related gene sequence SEQ ID NO:7603.
XX	
KW	Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW	stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW	cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
XX	gene; ds.
OS	Homo sapiens.
XX	
FN	W0200194629-A2.
XX	
PD	13-DEC-2001.
XX	
PF	30-MAY-2001; 2001WQ-US10838.
XX	
PR	05-JUN-2000; 2000US-209473P.
PR	05-JUN-2000; 2000US-209531P.
PR	18-SEP-2000; 2000US-23133P.
PR	18-SEP-2000; 2000US-231617P.
PR	20-SEP-2000; 2000US-234009P.
PR	20-SEP-2000; 2000US-234034P.
PR	20-SEP-2000; 2000US-234052P.
PR	22-SEP-2000; 2000US-234509P.
PR	22-SEP-2000; 2000US-234567P.
PR	25-SEP-2000; 2000US-234923P.
PR	25-SEP-2000; 2000US-234924P.
PR	25-SEP-2000; 2000US-235077P.
PR	25-SEP-2000; 2000US-235082P.
PR	25-SEP-2000; 2000US-235134P.
PR	25-SEP-2000; 2000US-235280P.
PR	26-SEP-2000; 2000US-235637P.
PR	26-SEP-2000; 2000US-235638P.
PR	27-SEP-2000; 2000US-235711P.
PR	27-SEP-2000; 2000US-235720P.
PR	27-SEP-2000; 2000US-235840P.
PR	27-SEP-2000; 2000US-235863P.
PR	28-SEP-2000; 2000US-236032P.
PR	28-SEP-2000; 2000US-236033P.
PR	28-SEP-2000; 2000US-236034P.
PR	28-SEP-2000; 2000US-236109P.
PR	28-SEP-2000; 2000US-236111P.
PR	29-SEP-2000; 2000US-236842P.
PR	29-SEP-2000; 2000US-236891P.
PR	02-OCT-2000; 2000US-237172P.
PR	02-OCT-2000; 2000US-237173P.
PR	02-OCT-2000; 2000US-237278P.
PR	02-OCT-2000; 2000US-237294P.
PR	02-OCT-2000; 2000US-237295P.
PR	02-OCT-2000; 2000US-237316P.
PR	03-OCT-2000; 2000US-237425P.
PR	03-OCT-2000; 2000US-237589P.
PR	03-OCT-2000; 2000US-237604P.
PR	03-OCT-2000; 2000US-237606P.
PR	03-OCT-2000; 2000US-237608P.
PR	01-NOV-2000; 2000US-244867P.
PR	01-NOV-2000; 2000US-245084P.
XX	
PA	(AVAL-) AVALON PHARM.
XX	
PI	Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;
PI	Sopet DR, Weaver Z;
XX	
DR	WPI; 2002-188264/24.
XX	
PT	Screening for anti-neoplastic agent involves exposing cells to a chemical agent to be tested for anti-neoplastic activity, and











196 GGCAGCAACATCATCAGCAACCTCTTCAGAGAGAGCTGCCAGAGATCATGATACCTC 137  
562 TTCTCGGAGAGCTGACCTGATTTGAAATGACGCCATTGTGTAGCTGATATGATC 621  
136 TTCTCGGAGAGCTGACCTGATTTGAAATGACGCCATTGTGTAGCTGATATGATC 77  
622 TTGAGATGATTTGAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 681  
76 TTGAGATGATTTGAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 17  
682 G 682  
16 G 16

T 9  
432/c  
AA68432 standard; cDNA; 311 BP.  
AA68432;

12-APR-2001 (first entry)  
Human lung tumour protein related nucleotide sequence SEQ ID NO:367.

Human; lung cancer; lung tumour; lung tumour protein; gene therapy;  
lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;  
cytostatic; antisense inhibition; ss.

Homo sapiens.  
MO200100828-A2.

04-JAN-2001.  
30-JUN-2000; 2000MO-US18061.

30-JUN-1999; 99US-0346492.  
15-OCT-1999; 99US-0419356.  
10-DEC-1999; 99US-0468867.  
30-DEC-1999; 99US-0476300.  
06-MAR-2000; 2000US-0519642.  
22-MAR-2000; 2000US-0531077.  
10-APR-2000; 2000US-0546259.  
27-APR-2000; 2000US-0560406.  
05-JUN-2000; 2000US-0589184.

(CORI-) CORIXA CORP.  
Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
Rettler MW, Mannion J;  
#PI; 2001-071488/08.

lung tumor-associated proteins and the nucleic acids that encode them,  
useful for preventing, diagnosing and treating lung cancer -  
Claim 4; Page 280; 436pp; English.

The present invention describes immunogenic portions of lung tumour-  
associated proteins (I) and the nucleic acids (NAs) that encode them.  
(I) have cytostatic activity and can be used in gene therapy, antisense  
inhibition and in vaccines. The NAs and the lung tumour-associated  
proteins they encode may be used in the prevention, treatment and  
diagnosis of diseases associated with their inappropriate expression,  
especially lung cancers. For example, the NAs may be administered to  
treat diseases by rectifying mutations or deletions in a patient's genome  
that affect the activity of the proteins by expressing inactive proteins  
or to supplement the patients own production of (I). Additionally, the  
NAs may be used to produce the lung-tumour associated protein, according  
to standard recombinant DNA methodology. Conversely, antisense NA  
molecules may be administered to down regulate protein expression by

CC binding with the cells own genes and preventing their expression. The NA  
CC and complementary sequences may also be used as DNA probes in diagnostic  
CC assays to detect and quantitate the presence of similar NA sequences in  
CC samples, and hence which patients may be in need of treatment for lung  
CC cancer. The (I) may be used as antigens in the production of antibodies  
CC and in assays to identify modulators (agonists and antagonists) of the  
CC expression and activity of the protein. AA68083 to AA68878 and  
CC AA67684 to AA67687 represent human lung tumour protein related  
CC nucleotide and protein sequences which are used in the exemplification  
CC of the present invention.

XX Sequence 311 BP; 70 A; 81 C; 94 G; 66 T; 0 other;  
SQ

Query Match 31.3%; Score 213.4; DB 22; Length 311;  
Best Local Similarity 82.7%; Pred. No. 5e-54;  
Matches 244; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 388 GCCAACAATGCCAAGCTGTGTGAAGACTTTCATGAGACCTCACTGTTGTGCTCC 447  
DB 310 GCCAACAATGCCAAGCTGTGTGAAGACTTTCATGAGACCTCACTGTTGTGCTCC 251  
QY 448 AACGCACTGACCACTGACTACCACTGACTGAGAAAGAGCTGTGCTCCAGGCGGC 507  
DB 250 AGCAGACTGACCTGTTTGAACCTCACTGACTGAGAAAGAGCTGTGCTCCAGGCGGC 191  
QY 508 AACATCTACCCCTTACTGAGAGAGATTGTCATGAGAAATGATGAGCTCTTCT 567  
DB 190 AACATCTACCAACCTCTTCAAGAGAGCTGCCACAGAGATGATGAGCTCTTCTCC 131  
QY 568 GGGAGCTGTACCTGATGAGAAATGACCACTGTTGTGCTGATATGATCTTTGAG 627  
DB 130 GGGAGCTGTACCTGATGAGAAATGACCACTGTTGTGCTGATATGATCTTTGAG 71  
QY 628 ATGATCTGAGCAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 682  
DB 70 ATGATCTGAGCAATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 16

RESULT 10  
ABK38343/c  
ID ABK38343 standard; cDNA; 311 BP.  
XX ABK38343;  
AC 21-MAY-2002 (first entry)  
XX 21-MAY-2002 (first entry)  
XX CDNA encoding clone #26488 of lung tumour protein.  
DE CDNA encoding clone #26488 of lung tumour protein.  
XX Lung tumour; cancer; T cell; immune response stimulator;  
XX cytostatic; gene; ss.  
XX Homo sapiens.  
XX MO200204514-A2.  
XX 17-JAN-2002.  
XX 10-JUL-2001; 2001WO-US22058.  
XX 11-JUN-2000; 2000US-0614124.  
XX 29-AUG-2000; 2000US-0615633.  
XX 08-SEP-2000; 2000US-0658824.  
XX 26-SEP-2000; 2000US-0671325.  
XX 06-OCT-2000; 2000US-0677419.  
XX 30-OCT-2000; 2000US-0702705.  
XX 13-DEC-2000; 2000US-0736457.  
XX 03-MAY-2001; 2001US-0849626.  
XX (CORI-) CORIXA CORP.  
XX Wang T, Matenabe Y, Henderson RA, Johnson JC, Rettler MW,  
XX Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS;  
XX Mcnabb A, Wang A, Fanger N, Switzer A, McNeill PD, Clapper JD;

PI: 2002-164634/21.

ovel polynucleotide encoding a lung tumour polypeptide useful for stimulating and/or expanding T cells specific for a tumour protein - sample 1; SEQ ID No 367; 223pp; English.

he invention describes an isolated polynucleotide and polypeptide useful for stimulating and/or expanding T cells specific for a tumour protein for determining the presence of a cancer in a patient. A tumour composition containing the polynucleotide and/or polypeptide is useful for treating a lung cancer in a patient. The polypeptide is useful for moving tumour cells from a biological sample. The polynucleotide is also useful as probe or primer to detect the level of mRNA encoding a tumour protein. This sequence encodes a lung tumour associated protein X protein fragment, described in the method of the invention.

te: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

sequence 311 BP; 70 A; 81 C; 94 G; 66 T; 0 other;

Y Match 31.3%; Score 213.4; DB 24; Length 311; Local Similarity 82.7%; Pred. No. 5e-54; Mismatches 51; Indels 0; Gaps 0;

388 GCCAACAAATGCCAAGCTGTGTGAGAACTTTTCATAGACCTCAACTGTGTGCTCC 447  
|||||  
310 GCCAACAAATGCCAAGCTGTGTGAGAACTTTTCATAGACCTCAACTGTGTGCTCC 251  
|||||  
448 AACGCACTGACCACTGACTACCACTGAGAGAAAGCTGTGTGCTGAGGCGC 507  
|||||  
250 AGCAGACTGACTGTGTTGACCACTGAGTCTGAGAGAACTTTGTCTCGGCGAC 191  
|||||  
508 AACGCACTGACCACTGACTACCACTGAGAGAAAGCTGTGTGCTGAGGCTCT 567  
|||||  
190 AACGCACTGACCACTGACTACCACTGAGAGAAAGCTGTGTGCTGAGGCTCT 131  
|||||  
568 AACGCACTGACCACTGACTACCACTGAGAGAAAGCTGTGTGCTGAGGCTCT 627  
|||||  
130 AACGCACTGACCACTGACTACCACTGAGAGAAAGCTGTGTGCTGAGGCTCT 71  
|||||  
628 AACGCACTGACCACTGACTACCACTGAGAGAAAGCTGTGTGCTGAGGCTCT 682  
|||||  
70 AACGCACTGACCACTGACTACCACTGAGAGAAAGCTGTGTGCTGAGGCTCT 16  
|||||

11  
72/c  
2A10672 standard; cDNA; 311 BP.  
2A10672;

5-JUN-2003 (first entry)

uman lung neuroendocrine carcinoma library MLN1 cDNA, SEQ ID 367.

uman lung cancer; ss; lung tumour; cytostatic; vaccine;  
cell expansion; CD4; CD8.

omo sapiens.

32002197669-A1.

5-DEC-2002.

5-MAY-2001; 2001US-0849626.

5-DEC-2000; 2000US-0736457.

WANG/) BANGUR C S.

PA (FANG/) FANGER G R.  
PA (WANG/) WANG A.  
PA (WANG/) WANG T.  
PA (SMIT/) SWITZER A P.  
PA (MCNE/) MCNEILL P D.  
PA (CLAF/) CLAPPER J D.  
XX  
XX  
PI Bangur CS, Fanger GR, Wang A, Wang T, Switzer AP, McNeill PD;  
PI Clapper JD;  
XX  
XX  
DR WPI; 2003-352750/33.

XX  
XX  
PT Novel lung cancer polynucleotide encoding lung cancer protein, useful  
PT for detecting the presence of lung cancer in a patient, and in  
PT pharmaceutical compositions, e.g. vaccines, for treating lung cancer -  
PS Example 1; Page -; 72pp; English.

XX  
XX  
CC The invention relates to a polynucleotide encoding a lung tumour protein,  
CC comprising a sequence selected from any of the 14 sequences  
CC mentioned in the specification, or a sequence (82) mentioned in  
CC specification, complement of S1, sequences consisting of at least 20  
CC contiguous residues of S1, sequences that hybridise to S1, sequences  
CC having 75%, preferably 90%, identity to S1, or degenerate variants of  
CC S1. Also included are an isolated polypeptide (comprising a sequence (53)  
CC selected from any one of the 4 amino acid sequences mentioned in the  
CC specification, a sequence encoded by the polynucleotide, or sequences  
CC having at least 70%, preferably 90%, identity to a sequence encoded by  
CC the polynucleotide), an expression vector comprising the polynucleotide  
CC operably linked to an expression control sequence, a host cell  
CC transformed or transfected with the vector, an isolated antibody (or its  
CC antigen-binding fragment) that specifically binds to the polypeptide,  
CC detecting the presence of a cancer in a patient, a fusion protein  
CC comprising the polypeptide, an oligonucleotide that hybridises to  
CC S1 under moderately stringent conditions, stimulating and/or expanding T  
CC cells specific for a tumour protein (comprising contacting T cells with  
CC the polynucleotide, protein or antigen-presenting cells, under conditions  
CC and for a time sufficient to permit the stimulation and/or expansion of T  
CC cells) and inhibiting the development of a cancer in a patient (by the  
CC incubating CD4<sup>+</sup> and/or CD8<sup>+</sup> T cells isolated from a patient with the  
CC polynucleotide, protein or antigen-presenting cells that express the  
CC polynucleotide, such that T cells proliferate, administering to the  
CC patient an effective amount of the proliferated T cells, and thus  
CC inhibiting the development of a cancer in the patient. The  
CC polynucleotide, protein and cells are useful in a composition for  
CC stimulating an immune response in a patient, and for treating a cancer in  
CC a patient (particularly lung cancer). The oligonucleotide is useful for  
CC determining the presence of a cancer in a patient. The protein and  
CC oligonucleotides are useful in pharmaceutical compositions, e.g.  
CC vaccines. The polynucleotide is also useful as a probe or primer for  
CC nucleic acid hybridisation, and in the design and preparation of  
CC ribozyme molecules for inhibiting expression of tumour polypeptides and  
CC proteins in tumour cells. An amplified portion of the polynucleotide is  
CC useful for isolating a full-length gene from a suitable library.  
CC The present sequence is a cDNA (full length, extended or partial)  
CC isolated from a library derived from lung tumour/cancer cells.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from the USPTO  
CC at [seqdata.uspto.gov/sequence.html?DocId=20020197669](http://seqdata.uspto.gov/sequence.html?DocId=20020197669).  
XX  
XX  
SQ Sequence 311 BP; 70 A; 81 C; 94 G; 66 T; 0 other;

Query Match 31.3%; Score 213.4; DB 25; Length 311;  
Best Local Similarity 82.7%; Pred. No. 5e-54;  
Matches 244; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 388 GCCAACAAATGCCAAGCTGTGTGAGAACTTTTCATAGACCTCAACTGTGTGCTCC 447  
|||||  
DB 310 GCCAACAAATGCCAAGCTGTGTGAGAACTTTTCATAGACCTCAACTGTGTGCTCC 251  
|||||  
OY 448 AACGCACTGACCACTGACTACCACTGAGAGAAAGCTGTGTGCTGAGGCGC 507  
|||||

250 AGCACTGACCTGCTTGGACCACTCAGTGCACGAACATTGTGTCCCTCGGGCAGC 191  
 508 AACATCTGACCCCTTACTGACGAGATTTGTCAATGAGAAATGAGACTCTTCT 567  
 190 AACATCTGACCAACCTCTTCAAGAGAGAGACCAACGAAGATCGATGACCTCTCC 131  
 568 GGAAGCTGTACTCTCATTTGAAATTCAGCCATTGTGGTAGCTCATTAATCTTTGAG 627  
 130 GGAAGCTGTACTCTCATTTGAAATTCAGCCATTGTGGTAGCTCATTAATCTTTGAG 71  
 628 ATGATTTGAGCATGT 682  
 70 ATGATTTGAGCATGT 16  
 T 12  
 623/C  
 ABX99623 standard; cDNA; 311 BP.  
 ABX99623;  
 22-MAY-2003 (first entry)  
 Lung cancer therapy and diagnosis associated cDNA #350.  
 Lung cancer; cytostatic; vaccine; gene therapy; cancer;  
 gene; ss.  
 Homo sapiens.  
 US2002172952-A1.  
 21-NOV-2002.  
 10-JUL-2001; 2001US-0902941.  
 30-JUN-1999; 99US-0346492.  
 15-OCT-1999; 99US-0419356.  
 17-DEC-1999; 99US-0466867.  
 30-DEC-1999; 99US-0476300.  
 06-MAR-2000; 2000US-0519642.  
 22-MAR-2000; 2000US-0533077.  
 10-APR-2000; 2000US-0546259.  
 27-APR-2000; 2000US-0560406.  
 05-JUN-2000; 2000US-0589184.  
 11-JUL-2000; 2000US-0614124.  
 29-AUG-2000; 2000US-0651563.  
 08-SEP-2000; 2000US-0658824.  
 26-SEP-2000; 2000US-0671325.  
 06-OCT-2000; 2000US-0677419.  
 30-OCT-2000; 2000US-0702705.  
 13-DEC-2000; 2000US-0736457.  
 03-MAY-2001; 2001US-0849626.  
 (CORI-) CORIXA CORP.  
 Henderson RA, Wang T, Matanabe Y, Johnson JC, Retter MW, Durham M;  
 Carter D, Ranger GR, Vedvick TS, Bangur CS, McNabb A;  
 FBI; 2003-328427/31.  
 new polynucleotide, useful for preparing a composition for treating or  
 inhibiting development of cancer, e.g. lung cancer -  
 Example 1; SEQ ID NO 367; 82bp; English.

The invention describes an isolated polynucleotide comprising one of 32  
 sequences, complement or degenerate variants of them. The polynucleotide  
 is useful for preparing a composition e.g. a vaccine or for gene therapy,  
 for treating or inhibiting development of cancer, e.g. lung cancer.  
 This sequence represents a polynucleotide associated with the  
 compositions and methods for the therapy and diagnosis of lung cancer.

SO Sequence 311 BP; 70 A; 81 C; 94 G; 66 T; 0 other;  
 Query Match 31.3%; Score 213.4; DB 25; Length 311;  
 Best Local Similarity 82.7%; Pred. No. 5e-54;  
 Matches 244; Conservative 0; Mismatches 51; Indels 0; Gaps 0;  
 QY 388 GCCAACAATCCCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 447  
 DB 310 GCCAACAAGCCCAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 251  
 QY 448 AACGACCTGACCACTGTACTGACCACTGACCACTGACCACTGACCACTGACCACTG 507  
 DB 250 AGCACTGACCTGCTTGT 191  
 QY 508 AACATCTGACCCCTTACTGACGAGATTTGTCAATGAGAAATGAGACTCTTCTCT 567  
 DB 190 AACATCTGACCAACCTCTTCAAGAGAGAGACTGACCAAGAGATCGATGACCTCTCC 131  
 QY 568 GGAAGCTGTACTCTCATTTGAAATTCAGCCATTGTGGTAGCTCATTAATCTTTGAG 627  
 DB 130 GGAAGCTGTACTCTCATTTGAAATTCAGCCATTGTGGTAGCTCATTAATCTTTGAG 71  
 QY 628 ATGATTTGAGCATGT 682  
 DB 70 ATGATTTGAGCATGT 16

RESULT 13  
 AAX37230  
 ID AAX37230 standard; DNA; 357 BP.  
 AC AAX37230;  
 XX  
 DT 07-JUL-1999 (first entry)  
 DE DNA sequence of EC2 fragment cloned in pThio-His C.  
 XX  
 KW CD81 protein; receptor; Hepatitis C virus; HCV; E2 protein; infection;  
 KM immunogen; therapy; diagnosis; EC2 region; ss.  
 OS Homo sapiens.  
 PN WO9918198-A1.  
 XX  
 PD 15-APR-1999.  
 XX  
 PF 06-OCT-1998; 98WO-1B01628.  
 XX  
 PR 23-JUN-1998; 98GB-0013560.  
 XX  
 PR 06-OCT-1997; 97GB-0021182.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 XX  
 PI Abirignani S, Grandi G;  
 XX  
 DR WPI; 1999-264018/22.  
 DR P-PSDB; AAY09154.  
 XX  
 PT Hepatitis C receptor protein CD81  
 XX  
 PS Example 6; Fig 5; 55bp; English.  
 CC  
 CC The invention relates to the use of the CD81 protein that is identified  
 CC as a receptor for Hepatitis C virus (HCV). Compounds that specifically  
 CC bind to CD81 can be used for therapy and diagnosis of HCV. The CD81  
 CC protein binds to HCV, preferably to the E2 protein. The CD81 protein can  
 CC be used to raise antibodies, which can then be used in the diagnosis or  
 CC therapy of HCV. The antibodies can be used to prevent the virus binding  
 CC to patient cells and being internalized. The CD81 protein can also be  
 CC used to treat HCV infection. The CD81 protein, a functional equivalent,  
 CC or a compound that binds to CD81, can be used in the manufacture of a  
 CC medicament for the treatment and diagnosis of a HCV infection. The CD81  
 CC protein can also be used as a protective immunogen in the control of

CV. The present sequence represents a DNA coding a EC2 fragment. EC2 is the region in CD81 that can bind to the E2 protein of HCV.

sequence 357 BP; 87 A; 101 C; 96 G; 73 T; 0 other;

Y Match 28.6%; Score 195.2; DB 20; Length 357;

Local Similarity 81.1%; Pred. No. 1.6e-48; Indels 0; Gaps 0;

hes 227; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

298 ATCTGGGGCTTGTAAACAAAGACGATCGCAAGATGTGAAGACCTTATGACGAG 357

70 AGCTGAGAGCTTGTCAACAGAGACGATCGCAAGATGTGAAGACCTTATGACGAG 129

358 GCCCTTACGAGCTGTATGATGATATGCAACATGCCAAGCTGTGTGAAGACT 417

130 GCCCTTACGAGCTGTATGATGATATGCAACATGCCAAGCTGTGTGAAGACT 189

418 TTCCATGAGAGCTTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 477

190 TTCCATGAGAGCTTCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 249

478 CTGAGAACAGCTGT 537

250 CTGAGAACAGCTGT 309

538 TGTGATCAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 577

310 TGCCACGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 349

14

31 AX37231 standard; DNA; 348 BP.

AX37231;

7-JUL-1999 (first entry)

NA sequence of EC2-His6 fragment cloned into pGEX-KG.

D81 protein; receptor; Hepatitis C virus; HCV; E2 protein; infection;

immunogen; therapy; diagnosis; EC2 region; ss.

omo sapiens.

09918198-A1.

5-APR-1999.

6-OCT-1998; 98WO-1B01628.

3-JUN-1998; 98GB-0013560.

6-OCT-1997; 97GB-0021182.

CHIR-) CHIRON SPA.

brlgnani S, Grandi G;

PI; 1999-264018/22.

-PSDB; AAY09155.

epatitis C receptor protein CD81

xample 6; Fig 8; 55pp; English.

he invention relates to the use of the CD81 protein that is identified

s a receptor for Hepatitis C virus (HCV). Compounds that specifically

ind to CD81 can be used for therapy and diagnosis of HCV. The CD81

rotein binds to HCV, preferably to the E2 protein. The CD81 protein can

be used to raise antibodies, which can then be used in the diagnosis or

therapy of HCV. The antibodies can be used to prevent the virus binding

o patient cells and being internalized. The CD81 protein can also be

sed to treat HCV infection. The CD81 protein, a functional equivalent,

CC or a compound that binds to CD81, can be used in the manufacture of a

CC medicant for the treatment and diagnosis of a HCV infection. The CD81

CC protein can also be used as a protective immunogen in the control of

CC HCV. The present sequence represents a DNA coding a EC2 fragment. EC2 is

the region in CD81 that can bind to the E2 protein of HCV.

XX Sequence 348 BP; 84 A; 96 C; 90 G; 78 T; 0 other;

SO Query Match 28.0%; Score 190.8; DB 20; Length 348;

Best Local Similarity 81.0%; Pred. No. 3.4e-47;

Matches 222; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 307 TTGTTAAACAAAGACGATCGCAAGATGTGAAGACCTTATGACGAGCCCTTACG 366

Db 52 TTGTTAAACAAAGACGATCGCAAGATGTGAAGACCTTATGACGAGCCCTTACG 111

QY 367 CAAGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 426

Db 112 CAGGCGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 171

QY 427 ACCCTCAACTGT 486

Db 172 ACCCTCAACTGT 231

QY 487 AGCTGTGTCTTCAAGCGGCAACATGATGATGATGATGATGATGATGATGATGAT 546

Db 232 AATTGTGTCTTCAAGCGGCAACATGATGATGATGATGATGATGATGATGATGAT 291

QY 547 AAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 580

Db 292 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 325

RESULT 15

AAQ29182

ID AAQ29182 standard; DNA; 1120 BP.

XX AAQ29182;

AC 25-MAR-2003 (updated)

DT 05-MAR-1993 (first entry)

XX DNA encoding metastasis controlling peptide.

XX M31-15 monoclonal antibody; cancer cell motility; prophylaxis;

XX treatment; lung; glioblastoma; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

PH CDS 112..798

FT mat\_peptide 115..795

FT /tag= a

FT /tag= b

FT /note= "mature peptide"

XX BP508417-A2.

XX 14-OCT-1992.

XX 09-APR-1992; 92BP-0106093.

XX 12-APR-1991; 91JP-0079996.

XX 17-APR-1991; 91JP-0085396.

XX 07-FEB-1992; 92JP-0022321.

XX (TAKA) TAKEDA CHEM IND LTD.

XX (MIVA) MIYAKE M.

XX Ikeyama S, Koyama M, Miyake M, Senoo M;

XX WPI; 1992-341723/42.

XX P-PSDB; AAR7525.

New monoclonal antibody M31-15 specific for metastasis-controlling peptide - useful for treating and preventing cancer and metastasis

Claim 8; Fig 4; 34pp; English.

The sequence is that of DNA encoding a cancer cell surface protein which is capable of suppressing the motility of cancer cells. It is specifically recognised by the monoclonal (Mab) antibody M31-15 which is useful for suppressing cancer metastasis. The polypeptide encoded by the DNA and Mab M31-15 are therefore useful in the prophylaxis and treatment of cancer, e.g. lung, glioblastoma and other metastatic cancers.

(Updated on 25-MAR-2003 to correct FN field.)

Sequence 1120 BP; 250 A; 256 C; 278 G; 336 T; 0 other;

100% Match 22.1%; Score 151; DB 13; Length 1120;

Local Similarity 55.3%; Pred. No. 6.1e-35;

Conserved 373; Conservative 0; Mismatches 275; Indels 27; Gaps 3;

```
1 AAATACCTGCTCTGCTGCTCAATTTGCTCTGCTGCTGAGCGGTGATCTTAGGT 60
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
142 AAATACCTGCTGCTGCTGCTCAATTTGCTCTGCTGCTGAGCGGTGATCTTAGGT 120
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
61 GTAGCTGCTGCTGCTGCTGCTCAATTTGCTCTGCTGCTGAGCGGTGATCTTAGGT 120
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
202 ATTGACTATGCTGCTGCTGCTCAATTTGCTCTGCTGCTGAGCGGTGATCTTAGGT 180
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
121 AACAAACCGGACCCCAACACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
262 AATTAATTAATTC-----CAGCTTCAACAGAGGTATTTCTGATCGAGCGCGCC 315
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
181 GATGATGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
316 CTATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 375
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
241 CTGAGGACGCTTCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
376 CTGAGGACGCTTCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 435
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
301 TGGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
436 TGGGATATTTCCCAAGAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 495
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
361 CTTCACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
496 TACAAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 549
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
421 CATGAGACGCTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
550 CACTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 609
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
481 AGGAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
610 TGGCCCAAGAG-----GAGTACTGGAACCTTCAACGCTGGAAGTCTGT 654
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
541 CATGAGAAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
655 CCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 714
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
601 GTGTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
715 GGCAATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 774
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
661 CGGAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 675
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
775 CGCAGGAACCGCGAG 789
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 2  
US-09-702-705-367/c  
; Sequence 367, Application US/09702705

```

nt No. 6504010
RAL INFORMATION:
LICANT: Wang, Tonglong
LICANT: Bangur, Chaitanya S.
LICANT: Lodes, Michael A.
LICANT: Fanger, Gary
LICANT: Vedvick, Tom
LICANT: Carter, Darrick
LICANT: Retter, Marc
LICANT: Mannion, Jane
LICANT: Fan, Liqun
LE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
E OF INVENTION: DIAGNOSIS OF LUNG CANCER
E REFERENCE: 210121.478C14
RENT APPLICATION NUMBER: US/09/702,705
RENT FILING DATE: 2000-10-30
BER OF SEQ ID NOS: 1833
TWARE: FASTSEQ for Windows Version 3.0
ID NO 367
NGTH: 311
PE: DNA
GANISM: Homo sapien
702-705-367

Y Match 31.3%; Score 213.4; DB 4; Length 311;
Local Similarity 82.7%; Pred. No. 6.9e-55;
hes 244; Conservative 0; Mismatches 51; Indels 0; Gaps 0

388 GCCAACAATGCCAAGCTGTGTGTAAGACTTTCCATGAGAGCTCACTGTGTGGCTCC 447
310 GCCAACAAGCCAAAGCTGTGTGTAAGACTTTCCACGAGAGCTTGACTGTGTGGCTCC 251
448 AACGCACTGACCACTGACTACCAACCAATTACTGAGAGACAGCTGTGTCCCTCAGAGCGC 507
250 AGACACTGACTGTGCTTTGACCACTCACTGCTCAAGAACATTTGTGTCTCGGGCAGC 191
508 AACATTACTCACTCCCTTAAGTCTGAGCAAGATTGTATCGAAATGCAATGACCTTCTCT 567
190 AACATTAATGAGAACCTCTTCACAGAGAGACTGCCACACGAAGATGCAATGACTTCTCC 131
568 GGAAGCTTACCTCAATGGAATTTGACAGCAATTTGGAGAGCTGTGCAATTATGATTTGAG 627
130 GGAAGCTTACCTCAATGGAATTTGACAGCAATTTGGAGAGCTGTGCAATTATGATTTGAG 71
628 ATGATTTGAGCAGTGTGTGTGTGCTGTGTGTGCGAATCCGGAACAGCTCCGTGTACTGAG 682
70 ATGATTTGAGCAGTGTGTGTGTGCTGTGTGTGCGAATCCGGAACAGCTCCGTGTACTGAG 16

736-457-367/c
ence 367, Application US/09736457
ral No. 6509448
RAL INFORMATION:
LICANT: Wang, Tonglong
LICANT: Bangur, Chaitanya S.
LICANT: Lodes, Michael A.
LICANT: Fanger, Gary
LICANT: Vedvick, Tom
LICANT: Carter, Darrick
LICANT: Retter, Marc
LICANT: Mannion, Jane
LICANT: Fan, Liqun
LICANT: Wang, Aijun
LE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
E OF INVENTION: DIAGNOSIS OF LUNG CANCER
3 REFERENCE: 210121.478C15
RENT APPLICATION NUMBER: US/09/736,457
RENT FILING DATE: 2000-12-13
BER OF SEQ ID NOS: 1864
TWARE: FASTSEQ for Windows Version 3.0
ID NO 367
NGTH: 311

```

```

1      TYPE: DNA
2      ORGANISM: Homo sapien
3      US-09-736-457-367
4
5      Query Match      31.3%; Score 213.4; DB 4; Length 311;
6      Best Local Similarity 82.7%; Pred. No. 6.9e-55;
7      Matches 244; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
8
9      QY      GCCAACAAATGCCAAGGCTGTGTGGAAGACCTTTCCATGAGAGCGCTCAACTGTGTGGCTCC 447
10     |||||
11     DB      GCCAACAAAGCGCAAGGCTGTGTGGAAGACCTTTCCACAGAGCGCTGACTGTGTGGCTCC 251
12
13     QY      AACGCACTGACCAAGCTAGTACATACCAACATCTGTAGAGAAAGCGCTGTCTCCAGGCGAG 507
14     |||||
15     DB      AGCAGACATGACTGTGTTTGAACCACTCAGTGTCTAGAGAAATTTGTGTCTCCCGGGCAGC 191
16
17     QY      AACATCTACACCCCTTACTCTGACGACGAAGATTGTCAAGAAATGATGATGACTCTTCTT 567
18     |||||
19     DB      AACATCTACGCAACCTCTTCAAGAGAGACTGCGCACAGAGATGATGATGACTCTTCTCC 131
20
21     QY      GGAAGAGCTGATACCTCATTTGAAATTGACGACCATTTGTGTGTAAGCTGATGATGATGAG 627
22     |||||
23     DB      GGAAGAGCTGATACCTCATTTGACCATTTGCTGCAATGTGTGTGATGATGATGATGAG 71
24
25     QY      ATGATTTCTGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 682
26     |||||
27     DB      ATGATCTGTGACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 16
28
29 RESULT 4
30 US-08-254-493-2
31 Sequence 2, Application US/08254493
32 Patent No. 5439886
33 GENERAL INFORMATION:
34 APPLICANT: IKEYAMA, SHUTCHI
35 APPLICANT: KOYAMA, MASARU
36 APPLICANT: MIYAKE, MASAYUKI
37 APPLICANT: SENO, MASAHARU
38 TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND
39 TITLE OF INVENTION: PRODUCTION THEREOF
40 NUMBER OF SEQUENCES: 7
41 CORRESPONDENCE ADDRESS:
42 ADDRESSER: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS & CUSHMAN
43 STREET: 130 WATER STREET
44 CITY: BOSTON
45 STATE: MASSACHUSETTS
46 COUNTRY: US
47 ZIP: 02109
48 COMPUTER READABLE FORM:
49 MEDIUM TYPE: Floppy disk
50 COMPUTER: IBM PC compatible
51 OPERATING SYSTEM: PC-DOS/MS-DOS
52 SOFTWARE: Patent Release #1.0, Version #1.25
53 CURRENT APPLICATION DATA:
54 APPLICATION NUMBER: US/08/254,493
55 FILING DATE:
56 CLASSIFICATION: 435
57 PRIOR APPLICATION DATA:
58 APPLICATION NUMBER: US 07/865552
59 FILING DATE: 09-APR-1992
60 PRIOR APPLICATION DATA:
61 APPLICATION NUMBER: JP 079996-1991
62 FILING DATE: 12-APR-1991
63 PRIOR APPLICATION DATA:
64 APPLICATION NUMBER: JP 085396-1991
65 FILING DATE: 17-APR-1991
66 PRIOR APPLICATION DATA:
67 APPLICATION NUMBER: JP 022321-1992
68 FILING DATE: 07-FEB-1992
69 ATTORNEY/AGENT INFORMATION:
70 NAME: RESNICK, DAVID S.
71 REGISTRATION NUMBER: 34235
72 REFERENCE/DOCKET NUMBER: 41777

```



TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 523-3400  
TELEFAX: (617) 523-6440  
TELEX: 200291 STRB UR  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 687 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULAR TYPE: cDNA to genomic RNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: HUMAN  
CELL TYPE: BREAST CARCINOMA  
CELL LINE: ZR-75-1  
-254-493-2

ry Match 22.1%; Score 151; DB 1; Length 687;  
t Local Similarity 55.3%; Pred. No. 6.6e-36;  
ches 373; Conservative 0; Mismatches 275; Indels 27; Gaps 3;

```
1 AATACCTGCTCTTGTCTTCAATTTCTGCTGCTGCTGAGGCGTATCTAGGT 60
  |||||
31 AATACCTGCTGTTGGATTAACTTCTGCTGCTGCGGATGCTGCTCTTGGCC 90
  |||||
61 GATGCTGTGTTGGTGTATGATCCAGACACAGACCTGCTGATCTGGAACCTGGA 120
  |||||
91 ATTGACTATGCTGCTGATTCAGACTCTCAGACCAAGAGATCTTCAGAGAAATTAAT 150
  |||||
121 AACAAACCGGACCCGAAACCTTCTAAGGCGATCTAATCTCTGCTGAGAGCT 180
  |||||
151 AATATTAATTC-----CAGCTTCAACAGAGATCTAATTAATCTGATCGGACCGCC 204
  |||||
181 GTGATGATTTGTAGGCTTCTGCGGATCTAATGCGGATCCAGAGTCCAGTCTG 240
  |||||
205 CTCATGATCTGATGCTGCTTCTGCGGCTCTGCGGCTGCTGCGGATCCAGTGCATG 264
  |||||
241 CTGGGAGCTTCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
  |||||
265 CTGGGAGCTTCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 324
  |||||
301 TGGGCTTGTAAACAGACAGATGCGCAAGATGTAAGAGCTTCTATGACAGGCC 360
  |||||
325 TGGGATATTTCCCAAGAGATGAGATTAAGAGATGAGAGATTTTAAAGAGACCC 384
  |||||
361 CTTCAGCAGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
  |||||
385 TACACAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 438
  |||||
421 CATGAGAGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
  |||||
439 CACTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 498
  |||||
481 AGGAACAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
  |||||
499 TCCCCAGAGAG-----GACGATCTGAAACCTTCACCGTAAAGTCTGCT 543
  |||||
541 CATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
  |||||
544 CCTATGCTGATCAAAAGAGTCTTTCAGCAATTAATTCACATCATGCGCGAGTGGGATC 603
  |||||
601 GTGATGCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
  |||||
604 GGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 663
  |||||
661 CGGAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 675
  |||||
664 CGCAGAGAGCGGAG 678
```

US-08-408-222B-2  
Sequence 2, Application US/08408222B  
Patent No. 576727  
GENERAL INFORMATION:  
APPLICANT: Ikeyama, Shuichi  
APPLICANT: Koyama, Masaru  
APPLICANT: Miyake, Masayuki  
APPLICANT: Senoo, Masaharu  
TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Dike, Bronstein, Roberts & Cushman  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: PastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/408,222B  
FILING DATE: 22-MAR-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/254,493  
FILING DATE: 06-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP-079996-1991  
FILING DATE: 12-APR-1991  
APPLICATION NUMBER: JP-085396-1991  
FILING DATE: 14-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP-022321-1992  
FILING DATE: 07-FEB-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Resnick, David S.  
REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 41777-DIV  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400  
TELEFAX: 617-523-6440  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 687 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULAR TYPE: cDNA to genomic RNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM: human  
CELL TYPE: breast carcinoma  
CELL LINE: ZR-75-1  
US-08-408-222B-2

Query Match 22.1%; Score 151; DB 1; Length 687;  
Best Local Similarity 55.3%; Pred. No. 6.6e-36;  
Matches 373; Conservative 0; Mismatches 275; Indels 27; Gaps 3;

```
Oy 1 AATACCTGCTCTTGTCTTCAATTTCTGCTGCTGCTGAGGCGTATCTAGGT 60
  |||||
Db 31 AATACCTGCTGTTGGATTAACTTCTGCTGCTGCGGATGCTGCTCTTGGCC 90
  |||||
Oy 61 GATGCTGTGTTGGTGTATGATCCAGACACAGACCTGCTGATCTGGAACCTGGA 120
  |||||
Db 91 ATTGACTATGCTGCTGATTCAGACTCTCAGACCAAGAGATCTTCAGAGAAATTAAT 150
  |||||
```



541 CATCAGAAATGATGAGCTCTCTGCGAAGCTGTACCTGATTCGACATT 600  
 655 CCTGATGCATCAAGAGGCTCTTCGACATTAATTCACATCATCGCGCATC 714  
 601 GTGTAGCTGTCTATTAATGATCTTTGAGATGATTCGAGCATGCTGTGTCATC 660  
 715 GGCATTCGCGTGTGATGATATTTGGCATGATCTTCAGATGATCTGTGCTGCTATC 774  
 661 CGGACAGCTCCGCTG 675  
 775 CGCAGAACCGCGAG 789

3-254-493-3  
 Invention 3, Application US/08254493  
 Inventor No. 5439886

GENERAL INFORMATION:  
 APPLICANT: IKEYAMA, SHUICHI  
 APPLICANT: KOYAMA, MASARU  
 APPLICANT: MIYAKE, MASAYUKI  
 APPLICANT: SENO, MASAHARU  
 TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND  
 TITLE OF INVENTION: PRODUCTION THEREOF  
 NUMBER OF SEQUENCES: 7  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS & CUSHMAN  
 STREET: 110 WATER STREET  
 CITY: BOSTON  
 STATE: MASSACHUSETTS  
 COUNTRY: US  
 ZIP: 02109

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/254,493  
 FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/865552

FILING DATE: 09-APR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 079996-1991

FILING DATE: 12-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 085396-1991

FILING DATE: 17-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 022321-1992

FILING DATE: 07-FEB-1992

ATTORNEY/AGENT INFORMATION:

NAME: RESNICK, DAVID S.

REGISTRATION NUMBER: 34235

REFERENCE/DOCKET NUMBER: 41777

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 523-3400

TELEFAX: (617) 523-6440

TELEX: 200291 STR UR

FORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1120 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA to genomic RNA

HYPOHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: HUMAN

CELL TYPE: BREAST CARCINOMA  
 CELL LINE: ZR-75-1  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 112..795  
 FEATURE:  
 NAME/KEY: mat peptide  
 LOCATION: 115..795  
 US-08-254-493-3

Query Match 21.9%; Score 149.4; DB 1; Length 1120;  
 Best Local Similarity 55.1%; Pred. No. 2.6e-35;  
 Matches 372; Conservative 0; Mismatches 276; Indels 27; Gaps 3;

QY 1 AAATACCTGCTCTTCTGCTTCAATTCGATCTTCTGAGCTGAGGCGTATCTAGT 60  
 DB 142 AAATACCTGCTGTTGGATTTAATCTTCTGCTTGGCTTGGGATTTGCTGCTTCC 201  
 QY 61 GTAGCTGTGTGTGCTGATCCAGACCAAGCTGCTGTAAGTGGAACTGGGA 120  
 DB 202 ATTGACTATGCTCCGATTCGACTTCAGACCAAGGATCTTCAGCAAGAACTAAT 261  
 QY 121 AACAAACCGGACCCAGACCTTCTAGTGGGATCTAATCTTCAATTCGATCTGCTGAGCT 180  
 DB 262 AATATATATTC-----CAGCTTCTACACAGAGCTATATTCGATTCGAGCGGCGC 315  
 QY 181 GTATGATGTTTGTAGGCTTCTGAGGCTGATATGAGGCGCATCCAGAGTCCAGTGTCTG 240  
 DB 316 CTATGATGCTGTGTGCTTCTGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGCTGCTGAG 375  
 QY 241 CTGAGGAGCTTCTTCACTGCTGCTGATTCCTGTTGCTGTAAGTGGCTGAGGAGCT 300  
 DB 376 CTGAGGAGCTTCTTCACTGCTGCTGATTCCTGTTGCTGTAAGTGGCTGAGGAGCT 435  
 QY 301 TGGGAGCTTCTGTAACCAAGACCAAGATCCCAAGATGTAAGGAGCTTCTATGACAGGCC 360  
 DB 436 TGGGATATTCCTCAAGAGATGATGATTAAGAGAGTCCAGAGTTTTCAGAGAGCACC 495  
 QY 361 CTTCAGCAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
 DB 496 TACACAGAGCTGAAACCAAGATGAGCCCA-----GGGGAAGCTGAAAGCAGATC 549  
 QY 421 CATGAGAGCTCACTGTTGTGCTTCAAGCACTGACCACTGACTACACATATCTG 480  
 DB 550 CACTATGCGTTGAACTGCTGT 609  
 QY 481 AGGACAGCTGTGTGCTTCAAGGCGCAATCTACACCTTCTTCTGAGCAAGATTTG 540  
 DB 610 TCCCCAAGAG-----GACGTACTGAAACCTTCAAGCTGAGGATCTGCTGT 654  
 QY 541 CATCAGAAATGATGAGCTCTTCTGCGAAGCTGTACCTGATTCGATTCGATTCGATTCGATTC 600  
 DB 655 CCTGATGCATCAAGAGGCTCTTCGACATTAATTCACATCATTCGCGGAGTGGGCGATC 714  
 QY 601 GTGTAGCTGTCTATTAATGATCTTTGAGATGATTCGAGCATGCTGTGCTGCTGCTGCTGCT 660  
 DB 715 GGCATTCGCGTGTGATGATATTTGGCATGATCTTCAGATGATCTTGTGCTGTGCTATC 774  
 QY 661 CGGACAGCTCCGCTG 675  
 DB 775 CGCAGAACCGCGAG 789

RESULT 8  
 US-09-040-984-14  
 Sequence 14, Application US/09040984  
 Patent No. 6210883  
 GENERAL INFORMATION:  
 APPLICANT: Wang, Steven G.  
 APPLICANT: Wang, Tongtong  
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS  
 TITLE OF INVENTION: OF LUNG CANCER  
 NUMBER OF SEQUENCES: 86

RESPONSE ADDRESS:  
ADDRESS: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/040,994  
FILING DATE: 18-MAR-1998

CLASSIFICATION:  
TORTNEY/AGENT INFORMATION:  
NAME: Maki, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.456  
ELECTRONIC COMMUNICATION INFORMATION:  
TELEPHONE: 206-622-4900  
TELEFAX: 206-282-6031  
TELEX:

FORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 679 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
040-984-14

Y Match 15.3%; Score 104.4; DB 3; Length 679;  
Local Similarity 59.0%; Pred. No. 7.4e-22;  
hes 181; Conservative 0; Mismatches 120; Indels 6; Gaps 1;

1 AAATCCGCTCTGCTCTTCAATTTGCTTCTGCTGCTGAGGCGTAACTAGT 60  
|||||  
132 AAATCCGCTCTGCTCTTCAATTTGCTTCTGCTGCTGAGGCGTAACTAGT 191  
|||||

61 GTAGCTCTGTGCTGCTGCTATGATCCAGACCAAGCGCTGCTGATCTGAGG 120  
|||||  
192 ATTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 245  
|||||

121 AACAAACCGGCAACCAACCTTTCTAAGTGGGATCTTCAATTTGCTGAGGAGT 180  
|||||  
246 ACTAATNATNATNATTTTCAGCTTTCTACACAGGAGTCTAATTTCTGATGGATCCGGC 305  
|||||

181 GTGATGATTTTGTAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240  
|||||  
306 CTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 365  
|||||

241 CTGAGGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300  
|||||  
366 CTGAGGACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 425  
|||||

301 TGGGGCT 307  
|||||

426 TGGGGAT 432

9  
123-912-14  
ence 14, Application US/09123912A  
nt No. 6312695  
RAL INFORMATION:  
LICANT: Reed, Steven G.  
LICANT: Wang, Tongcong  
LB OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER  
E REFERENCE: 210121.455C1  
RENT APPLICATION NUMBER: US/09/123,912A  
RENT FILING DATE: 1998-07-27  
OR APPLICATION NUMBER: 09/040,802

PRIOR FILING DATE: 1998-03-18  
NUMBER OF SEQ ID NOS: 114  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 14  
LENGTH: 679  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: modified\_base  
LOCATION: (129)  
OTHER INFORMATION: where n is a, c, g or t  
LOCATION: (68)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (83)  
OTHER INFORMATION: where n is a, c, g or t  
LOCATION: (87)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (94)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (104)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (117)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (142)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (145)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (151)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (187)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (201)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (211)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (226)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (229)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (239)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (241)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (245)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (252)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (255)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (259)  
OTHER INFORMATION: where n is a, c, g or t  
NAME/KEY: modified\_base  
LOCATION: (303)

Query Match	15.3%;	Score 104.4;	DB 4;	Length 679;
Best Local Similarity	59.0%;	Pred. No. 7.4e-22;		
Matches 181; Conservative	0;	Mismatches 120;	Indels 6;	Gaps 1;

RESULT 10  
US-09-643-597-14  
; Sequence 14, Application US/09643597

Query Match	15.3%	Score 104.4;	DB 4;	Length 679;
Best Local Similarity	59.0%;	Pred. No. 7.4e-22;		
Matches 181;	Conservative	0;	Mismatches 120;	Indels 6;
				Gaps 1.

```

11
480-884A-14
ence 14, Application US/09480884A
nt No. 5482597
RL INFORMATION:
LICANT: Wang, Tongtong
LICANT: Pan, Liqun
LICANT: Hosken, Nancy A.
LICANT: Kalos, Michael D.
LICANT: Panger, Gary R.
LE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
LB OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
B REFERENCE: 210121.455C6
RENT APPLICATION NUMBER: US/09/480,884A
RENT FILING DATE: 2001-08-27
BER OF SEQ ID NOS: 330
TMAR: FastSeq for Windows Version 3.0
ID NO 14
NTH: 679
PR: DNA
GANISM: Homo sapien
ATURE:
ME/KEY: misc_feature
CATION: (1)...(679)
HER INFORMATION: n = A,T,C or G
480-884A-14

```

```

11
480-884A-14
ence 14, Application US/09480884A
nt No. 5482597
RL INFORMATION:
LICANT: Wang, Tongtong
LICANT: Pan, Liqun
LICANT: Hosken, Nancy A.
LICANT: Kalos, Michael D.
LICANT: Panger, Gary R.
LE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
LB OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
B REFERENCE: 210121.455C6
RENT APPLICATION NUMBER: US/09/480,884A
RENT FILING DATE: 2001-08-27
BER OF SEQ ID NOS: 330
TMAR: FastSeq for Windows Version 3.0
ID NO 14
NTH: 679
PR: DNA
GANISM: Homo sapien
ATURE:
ME/KEY: misc_feature
CATION: (1)...(679)
HER INFORMATION: n = A,T,C or G
480-884A-14

```

```

Qy 241 CTGGGACGCTTCTTCACTGCTTGTGATCTGTTGCTGAGTGGCTGACAGCATC 300
Db 366 CTGGGACGCTTCTTCACTGCTTGTGATCTGTTGCTGAGTGGCTGACAGCATC 425
Qy 301 TGGGGCT 307
Db 426 TGGGGAT 432

```

```

RESULT 12
US-09-542-615A-14
; Sequence 14, Application US/09542615A
; Patent No. 6518256
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Pan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Panger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C8
; CURRENT FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 350
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 679
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURES:
; NAME/KEY: misc_feature
; LOCATION: (1)...(679)
; OTHER INFORMATION: n = A,T,C or G
US-09-542-615A-14

```

```

Query Match 15.3%; Score 104.4; DB 4; Length 679;
Best Local Similarity 59.0%; Pred. No. 7.4e-22;
Matches 181; Conservative 0; Mismatches 120; Indels 6; Gaps 1;
Qy 1 AAATACCTGCTCTTGTCTTCAATTTGCTTCTTGTGCTGCTGAGGCGTATCTAGGT 60
Db 132 AAATACCTGCTCTTGTCTTCAATTTGCTTCTTGTGCTGCTGAGGCGTATCTAGGT 191
Qy 61 GTAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
Db 192 ATTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245
Qy 121 AACAAACGGGACCCAAACCTTCTAAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
Db 246 ACTAATATATATATTTTTCAGCTTCTTCAACAGAGTCTAATTTCTGATCGATCCGCGC 305
Qy 181 GTGATGATGTTTGTAGGCTTCTGAGGCTGCTAATGAGGCTGCTAATGAGGCTGCT 240
Db 306 CTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 365
Qy 241 CTGGGACGCTTCTTCACTGCTTGTGATCTGTTGCTGAGTGGCTGACAGCATC 300
Db 366 CTGGGACGCTTCTTCACTGCTTGTGATCTGTTGCTGAGTGGCTGACAGCATC 425
Qy 301 TGGGGCT 307
Db 426 TGGGGAT 432

```

```

RESULT 13
US-09-606-421B-14
; Sequence 14, Application US/09606421B
; Patent No. 6531315
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Pan, Liqun

```









440 GTGATGATGATTTGGTGGCTTCTGGGCTGCTACGGGGGCATCCAGGAATCCCATGTGCTTG 499

241 CTGGGGAGCGTTCTTCACTGCTGCTTGATGATCTGTTTGCCTGTGAGTGGCTGCAAGGCATC 300

500 CTGGGGAGCGTTCTTCACTGCTGCTGCTGATCTGTTTGGCTGTGAGTGGCTGCGGCGCATC 559

301 TGGGGCTTGGTAAACAAAGACAGATGGCCCAAGATGTAAGAGCAAGTTCTATGACCAAGGCC 360

560 TGGGGCTTGTTCACACAGACCAAGTCCCAAGATGTAAGAGCAAGTTCTATGACCAAGGCC 619

361 CTTCAGCAAGCTGTGATGATGATGATGATGCCAACAATGCCAAGGCTGTGGTGAAGCATTC 420

620 CTACAGCAGGCGCGTGGTGGATGATGAGCGCCAAACCGCAAGGCTGTGGTGAAGCATTC 679

421 CATGAGAGCGCTCACTGTGTGTGGCTTCCAAAGCATGACCAACTGATCAACCATTAATG 480

680 CACAGAGCGCTGATCTGTGTGGCTTCCAGCACACTGATCTGTGTTGAACCACTCAATGCTC 739

481 AGGAACAGCGCTGTGCTCCTCAGCGGCGCAACATCACTCAACCCCTTAAGTACGACCAAGATTG 540

740 AAGAACATTTGTGTCTCCTCGGGGCAAGCAATCATGACGAACCTCTTCAAGAGAGATGCG 799

541 CATGAGAAATCGATGAGCTCTTCTCTGGGAAGCTGTAAGTCAATTTGAATTCAGCCATT 600

800 CACCAAGAGATCGATGAGCTCTCTCTCCGGGAAGCTGTAAGTCAATTCGAGCATTTGCGCATC 859

601 GTGGTACCTGTCAATTAAGATCTTTGAGATGATTCGAGACAGGAGTGGTGGTGGCATC 660

860 GTGGTCCCTGTGATCAATGATTTTCAGAGTATCTTGAGCAATGGTGTGTGGTGGCATC 919

661 CGGAACAGCTTCCGTGTAAGTGAAG 682

920 CGGAACAGCTTCCGTGTAAGTGAAG 941

2

969-708-132

ence 132, Application US/09969708

nt No. US20020102532A1

RAL INFORMATION:

LICANT: Augustus, Meana

LE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu

LE OF INVENTION: Sate

Qy	121	AAACAAACCGGACCCCAACACCTTCTTAAGTGGGCATCTACATCTTCATGCTGTGGAGCT	180
Db	389	GAACACCCCGCCCCAAACACTTCTATATGAGCAATCAATCCTCATTCGCTGTGGCGCT	448
Qy	181	GTGATGATGTTGTATGAGCTTCTGGGGGTATATGGGGCAATCCAGAGATCCCAATGTCTG	240
Db	449	GTATATATGTTGTTGGCTTCTTGGACTGTACGGGGCAATCCAGAAATCCCAATGTCTG	508
Qy	241	CTGGAGGACGTTCTTCAACCTGCTTGTGATCCCTGTTTGTCTGTGAGTGTGCTGCAGGATC	300
Db	509	CTGGGACAGTTTCTTCAACCTGCTGTGATCCCTGTTTGTCTGTGAGTGTGCTGCAGGATC	568
Qy	301	TGGGGCTTGTGTAAACAAAGACCAATTCGCAAGATGTGAAGCATTTATATACCAAGCC	360
Db	569	TGGGGCTTGTGTAAACAAAGACCAATTCGCAAGATGTGAAGCATTTATATACCAAGCC	628
Qy	361	CTTCAGACAAGCTGTATGGATGTATATCCCAACATGCCAGAGCGCTGTGGAGACATTTTC	420
Db	629	CTACAGCAAGCCGTGTGTATGTATATACCCCAACAGCCCAAGGCTGTGTGTAAACCTTC	688
Qy	421	CATGAGACGCTCAACTGTTGTGTGACTCCAAAGCACTGACCAACTGATACCAACATATCTG	480
Db	689	CACGAAACGCTTGACTGCTGTGTGCTCCAGCAACATGACTGTGTTGACCACTCAAGTCTC	748
Qy	481	AGGAAAGCGTGTGTGCTCCCAAGCGCGCAACATCTCAACCCCTTATCTGTACAGCAAAATGT	540
Db	749	AGGAAACAAATTTGTATGTCCCTCGGGCAGCAACATCAATGCAACCTCTTCAAGAGGACATGC	808
Qy	541	CATCAGAAATGATGAGACTCTTCTGTGGAAAGCTGTACCTATTTGAAATTTGACGACATT	600
Db	809	CACGAAATATGATGAGACTCTTCTCCGGAAAGCTGTATCTCATTCGCAATTTGCTGCATC	868
Qy	601	GTGTGACTGTCAATATATATCTTTGAGATGATTTGAGCAATGTGTGCTGTGTGATC	660
Db	869	GTGTGTGCTGTATATATATCTTTCAGATGATTCCTGAGCAATGTGTGCTGTGTGATC	928
Qy	661	CGGAACAGCTCGGTACTGAG	682
Db	929	CGGAACAGCTCGGTACTGAG	950

```

RESULT 3
US-10-171-581-120
: Sequence 120, Application US/10171581
: Publication No. US20030104426A1
: GENERAL INFORMATION:
: APPLICANT: Dai, Hongyue
: APPLICANT: Linaley, Peter
: APPLICANT: Mao Mao
: TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
: FILE REFERENCE: 9301-157-999
: CURRENT APPLICATION NUMBER: US/10/171,581
: PRIOR APPLICATION NUMBER: 60/236,914
: PRIOR FILING DATE: 2001-06-18
: NUMBER OF SEQ ID NOS: 366
: SEQ ID NO 120
: LENGTH: 1496
: TYPE: DNA
: ORGANISM: Homo sapiens
: PUBLICATION INFORMATION:
: DATABASE ACCESSION NUMBER: M33680
: DATABASE ENTRY DATE: 2001-06-18
: US-10-171-581-120

```

Query Match	78.4%	Score 534.8;	DB 15;	Length 1496;
Best Local Similarity	66.5%	Prod. No. 1.3e-165;		
Matches 599;	Conservative 0;	Mismatches 92;	Indels 0;	Gaps 0;

Qy	1	AAATACCTGCTCTTTCGTCATTTTCGTCCTTGCTGAGGCGATCTTAACT	60
Db	269	AAATACCTGCTCTTTCGTCATTTTCGTCCTTGCTGAGGCGATCTTAACT	328









```

APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darlick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Pan, Lijun
APPLICANT: Wang, Aijun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE REFERENCE: 210121.478C15
CURRENT APPLICATION NUMBER: US/09/736,457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
OPTABR: FastSeq for Windows Version 3.0
Q ID NO 367
LENGTH: 311
TYPE: DNA
ORGANISM: Homo sapien
9-736-457-367

ery Match 31.3%; Score 213.4; DB 10; Length 311;
at Local Similarity 82.7%; Pred. No.7.6e-60;
tches 244; Conservative 0; Mismatches 51; Indels 0; Gaps 0

388 GCCAACAATCCAAAGCGTGTGTGAAGACTTTCATAGAGCGTCAACGTGTGGCTCC 447
310 GCCAACAAGCCAAAGCGTGTGTGAAGACTTTCATAGAGCGTCAACGTGTGGCTCC 251
448 AAGCACTGACCAACTGACTACCACTACCACTAGAGAACGCGTGTGCTCCAGGCGGC 507
250 AGACACACTGACTCTTGTGACCACTCAGTGTCTCAAGAACATTGTGTGCTCCAGGCGGC 191
508 AACATCTGACCCCTTACTGACGAGAAATGTTCATCAGAAATGATAGTCTTTCTCT 567
190 AACATCTGACCAACTCTTCAAGAGAGACTGCAACAAAGATGATAGTCTTTCTCT 131
568 GCGAAGCTGACTCTCATTTGGAATTGAGCATGTGGTGTGCTCATATGATCTTTGAG 627
130 GCGAAGCTGACTCTCATTTGGAATTGAGCATGTGGTGTGCTCATATGATCTTTGAG 71
628 ATGATCTGAGCATGTGTGCTGTGCTGTGATCGGATCGGAAAGCTCCGTGATGAG 682
70 ATGATCTGAGCATGTGTGCTGTGCTGTGATCGGAAAGCTCCGTGATGAG 16

UT 13
9-902-941-367/c
quence 367, Application US/09902941
cent No. US20020172952A1
SERIAL INFORMATION:
APPLICANT: Henderson, Robert A.
APPLICANT: Wang, Tongtong
APPLICANT: Watanabe, Yoshihiro
APPLICANT: Johnson, Jeffrey C.
APPLICANT: Retter, Marc W.
APPLICANT: Marnerakis, Margarita
APPLICANT: Carter, Darlick
APPLICANT: Fanger, Gary R.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: McNabb, Andria
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE REFERENCE: 210121.478C17
CURRENT APPLICATION NUMBER: US/09/902,941
CURRENT FILING DATE: 2001-07-10
NUMBER OF SEQ ID NOS: 2002
OPTABR: FastSeq for Windows Version 4.0
LENGTH: 311

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-902-941-367

Query Match      31.3%; Score 213.4; DB 10; Length 311;
Best Local Similarity 82.7%; Pred. No. 7,66-60;
Matches 244; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 388 GCCAACATGCGCAGGCTGTGTGAAGACTTTCATGAGAGCGCTCAACTGTGTGCTCC 447
Db 310 GCCAACAAACGCCAAGGCTGTGTGAAGACTTTCACAGAGACGCTTGATGTGCTCC 251
OY 448 AACGACATGACCACTGACATCAACCATCTGAGAGAAACAGCTGTGTCTCCAGGCGGC 507
Db 250 AGCACTGATGTGTTTGAACCACTCAAGTGTCAAGAACTTTGTGTCTCCGCGGAGC 191
OY 508 AACATCTCAACCCCTTACTGACAGCAAGATTGTCAATCAGAAATCGATGAGCTCTCTC 567
Db 190 AACATCATCAGAACTCTTCAAGAGGAGACTGCACACAGAAATGATGATACCTCTTCC 131
OY 568 GCGAAGCTGTACTCTATTTGGAATTGACGCAATTTGTGTGTGCTGTCAATGATCTTTGAG 627
Db 130 GCGAAGCTGTACTCTATCGGATCTGCGCATCTGCGCATCGGAACAGCTCCGTACTGAG 71
OY 628 ATGATTTGAGCATTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 682
Db 70 ATGATCTGTAGCAATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 16

RESULT 14
US-09-849-626-367/c
; Sequence 367, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaifanya
; APPLICANT: Fanger, Gary
; APPLICANT: Mang, Aijun
; APPLICANT: Wang, Tonglong
; APPLICANT: Switzer, Anne
; APPLICANT: McNeill, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 367
; LENGTH: 311
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-849-626-367

Query Match      31.3%; Score 213.4; DB 10; Length 311;
Best Local Similarity 82.7%; Pred. No. 7,66-60;
Matches 244; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 388 GCCAACAAAGCGCAGGCTGTGTGAAGACTTTCATGAGAGCGCTCAACTGTGTGCTCC 447
Db 310 GCCAACAAAGCGCAGGCTGTGTGAAGACTTTCACAGAGCGCTTGATGTGCTCC 251
OY 448 AACGACATGACCACTGACATCAACCATCTGAGAGAAACAGCTGTGTCTCCAGGCGGC 507
Db 250 AGCAGCTGATGTGTTTGAACCACTCAAGTGTCAAGAACTTTGTGTCTCCGCGGAGC 191
OY 508 AACATCTCAACCCCTTACTGACAGCAAGATTGTCAATCAGAAATCGATGAGCTCTCTC 567
Db 190 AACATCATCAGAACTCTTCAAGAGGAGACTGCACACAGAAATGATGATACCTCTTCC 131
OY 568 GCGAAGCTGTACTCTATTTGGAATTGACGCAATTTGTGTGTGCTGTCAATGATCTTTGAG 627
Db 130 GCGAAGCTGTACTCTATCGGATCTGCGCATCTGCGCATCGGAACAGCTCCGTACTGAG 71

```

628 ATGATTTCTGAGCATGTGTCTGTCTGTGGCATCCGGAACAGCTCCGTACTTGAAG 682  
|||||  
70 ATGATCTTGAGCATGTGTCTGTCTGTGGCATCCGGAACAGCTCCGTACTTGAAG 16

T 15

-476-300-367/c

uence 367, Application US/09476300  
lication No. US20030125245A1

Publication No. US20030125245A1

**ERAL INFORMATION:**

PLICANT: Wang, Tongtong

PLICANT: Bangur, Chaitanya S.  
THE OR INVENTION. COMPOSTION

**TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND TREATMENT OF INFECTIONS**

FILE OF INVENTION: DIAGNOSIS OF LUNG CANCER

LE REFERENCE: 210121.478C3  
 PREPENT ADDITION NUMBER:

PRESENT FILING DATE: 1999-13-30

RENT FILING DATE: 1999-12-30

WEEK OF SEQ ID NOS: 785

FTWAVE: FastSeq for Windows Version 3.0  
ID NO 367

ID NO 367  
LENGTH 311

LENGTH: 311  
YPR: DM2

YPR: DNA  
RGANT SM.

REGANISM: Homo sapien  
-476-300-367

-416-300-361

ry Match	31.3%	Score 213.4;	DB 11;	Length 311;
Local Similarity	82.7%	Prod NO 7	50-60.	

Local Similarity 82.7%; Pred. No. 7.6e-60;  
Conservative 0; Mismatches 51; Indels 0; Gaps 0.

conservative 0; mismatches 51; indels 0; gaps 0;

388 GCGACCAATGCCAGGCTGTGTGAGAGCTTTCATGAGACGCTCAACTGTTGTGGCTCC 447

310 GCGAACACGCCAAGGCTGTGTGAAGACCTTCCACGAGACGCTTGACTGTGTGCTCC 251

448 AACGCACTGACCACTGACTACCACTACTGAGGAACAGCCTGTGTCCCTCAGGCGC 507

250 AGCACTGACTGCTTTGACCACTCAGTGCTCAAGACAAATTGTGTCCCTCGGCAGC 191

508 AACATACTCACCCTTACTGCAGCAGATTGTCATCAGAAATCGATGAGCTCTTCTCT 567

```

1 completed: December 9, 2003, 10:28:13
time : 1901 secs

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

nucleic - nucleic search, using sw model

on: December 9, 2003, 01:37:00 ; Search time 2618 Seconds  
(without alignments)  
6331.418 Million cell updates/sec

US-10-035-914-1

ect score: 682  
ence: 1 aaataccgctcctcgtctt.....gacacgctcgtctactgag 682

ing table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

ched: 22781392 seqs, 12152238056 residues

1 number of hits satisfying chosen parameters: 45562784

num DB seq length: 0  
num DB seq length: 2000000000

-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

base :

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estmu:\*
- 5: em\_estcov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hcc:\*
- 9: gb\_est1:\*
- 10: gb\_est2:\*
- 11: gb\_hcc:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: em\_gss\_hum:\*
- 18: em\_gss\_inv:\*
- 19: em\_gss\_pln:\*
- 20: em\_gss\_vit:\*
- 21: em\_gss\_fun:\*
- 22: em\_gss\_mam:\*
- 23: em\_gss\_mus:\*
- 24: em\_gss\_pro:\*
- 25: em\_gss\_rod:\*
- 26: em\_gss\_phg:\*
- 27: em\_gss\_vit:\*
- 28: gb\_gss1:\*
- 29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

lc	Query	Score	Match	Length	DB	ID	Description
1	682	100.0	885	12	BI695471	BI695471	BI695471 603346477
2	675.6	99.1	1169	13	BU515266	AGENCOURT	BU515266 AGENCOURT
3	675.4	99.0	918	13	BO900007	AGENCOURT	BO900007 AGENCOURT
4	661	96.9	881	14	CB196088	AGENCOURT	CB196088 AGENCOURT

5	660	96.8	698	12	BI157227	BI157227	BI157227 602923435
6	657.4	96.4	777	12	BI653596	AGENCOURT	BI653596 603300230
7	655.6	96.1	956	13	BO714590	AGENCOURT	BO714590 AGENCOURT
8	650.8	95.4	933	13	BU523591	AGENCOURT	BU523591 AGENCOURT
9	645	94.6	852	14	CB558103	AGENCOURT	CB558103 AGENCOURT
10	641.2	94.0	759	12	BI113178	AGENCOURT	BI113178 602900106
11	640	93.8	759	12	BI169388	AGENCOURT	BI169388 603345472
12	639.4	93.8	903	13	BO886993	AGENCOURT	BO886993 AGENCOURT
13	639	93.7	642	10	BF531191	AGENCOURT	BF531191 602073837
14	633.4	92.9	846	12	BI694804	AGENCOURT	BI694804 603347966
15	627.2	92.0	806	10	BF181041	AGENCOURT	BF181041 601805089
16	619	90.8	958	13	BO934764	AGENCOURT	BO934764 AGENCOURT
17	613.8	90.0	1002	13	BU539889	AGENCOURT	BU539889 AGENCOURT
18	612.4	89.8	647	10	BF138851	AGENCOURT	BF138851 601781760
19	611.4	89.6	945	13	BU147770	AGENCOURT	BU147770 AGENCOURT
20	607.2	89.0	918	10	BF139437	AGENCOURT	BF139437 601785421
21	605.8	88.8	884	13	BO886347	AGENCOURT	BO886347 AGENCOURT
22	603	88.4	867	13	BU505471	AGENCOURT	BU505471 AGENCOURT
23	601	88.1	1030	10	BF783079	AGENCOURT	BF783079 602109182
24	597	87.5	988	12	BI453448	AGENCOURT	BI453448 603170839
25	596	87.4	737	13	BI693368	AGENCOURT	BI693368 603343712
26	593	87.0	872	13	BU517509	AGENCOURT	BU517509 AGENCOURT
27	592.2	86.8	851	12	BI690870	AGENCOURT	BI690870 603312011
28	586.2	86.0	818	12	BI694417	AGENCOURT	BI694417 603346995
29	584	85.6	937	13	BO937206	AGENCOURT	BO937206 AGENCOURT
30	574.6	84.3	847	12	BI110858	AGENCOURT	BI110858 602895510
31	571.8	83.8	866	12	BI689533	AGENCOURT	BI689533 603315205
32	570	83.6	931	13	BO947699	AGENCOURT	BO947699 AGENCOURT
33	569	83.4	996	13	BO886352	AGENCOURT	BO886352 AGENCOURT
34	563.2	82.6	903	12	BI079904	AGENCOURT	BI079904 602875902
35	562.4	82.5	744	12	BI854539	AGENCOURT	BI854539 603380238
36	558.6	81.9	735	10	BG668442	DRABY07	BG668442 DRABY07
37	557	81.7	921	13	BO714029	AGENCOURT	BO714029 AGENCOURT
38	555.8	81.5	892	13	BO219107	AGENCOURT	BO219107 AGENCOURT
39	553.2	81.1	844	12	BI146915	AGENCOURT	BI146915 602911540
40	552.2	81.0	925	13	BO926886	AGENCOURT	BO926886 AGENCOURT
41	552	80.9	917	10	BF15104	AGENCOURT	BF15104 601780145
42	552	80.9	956	13	BU518267	AGENCOURT	BU518267 AGENCOURT
43	551.4	80.9	933	12	BI695295	AGENCOURT	BI695295 603345155
44	546.6	80.1	824	13	BU558495	AGENCOURT	BU558495 AGENCOURT
45	536.4	78.7	991	13	BU839845	AGENCOURT	BU839845 AGENCOURT

ALIGNMENTS

RESULT 1  
LOCUS BI695471  
DEFINITION 885 bp mRNA linear EST 18-SEP-2001  
603346477F1 NCI\_CGAP\_Mam2 Mus musculus cDNA clone IMAGE:5374171 5',  
mRNA sequence.  
ACCESSION BI695471  
VERSION BI695471.1 GI:15658100  
KEYWORDS  
SOURCE EST.  
ORGANISM Mus musculus (house mouse)  
MUS musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>.  
1 (bases 1 to 885)  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strauberg, Ph.D.  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
plate: L1AM1952 row: i column: 20  
High quality sequence stop: 816.



[illegible]

Tissue Procurement: Dr. David Rowe  
CDNA Library Preparation: Invitrogen Corp  
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
Plate: NDAM039 row: 1 column: 08  
High quality sequence start: 18  
High quality sequence stop: 71.

RES	Location/Qualifiers
source	1. .881

```

/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="TMAGS:30137839"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NH MGC 135"
/notes="Vector: pCMVSPORT6.1; Site 1: EcoRV; Site 2: NotI;
Normalized full-length enriched library from pooled mouse
embryonic limb, maxilla and mandible, day 12.5, 13.5, 14.5
and 15.5 (size selected for the 0.5-1 kb fragments)
Cloned directionally, priming method: Oligo-dT. cDNA
enrichment: >1k bp, Average insert size 1.6k bp.
Normalization (Cot value): 7.5 kb. Priming sequence:
5'GACCTGTTCAATGCGGAGGCGCCGCC(7)3' Tissue contributed by
David Rowe. Library constructed by Keegen, Invitrogen
Corp."

```

COUNT	174	211	231	265
N	a	c	g	t

```

Pair Match      96.9%; Score 661; DB 14; Length 881;
% Local Similarity 100.0%; Pred. No. 1.6e-164;
Ches 661; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

22 AATTTCGCTCTTCTGGCTGAGCGCTGAGCGGTATCCTAGAGTGTACTCTGAGGTGCGTAT 81  
27 AATTTCGCTCTTCTGGCTGAGCGGTATCCTAGAGTGTACTCTGAGGTGCGTAT 86  
82 GATTCACAGACCAACAGCTGTGTATACCTTGAAATCTGGAAACAAACCGACCCCAACACC 141  
87 GATTCACAGACCAACAGCTGTGTATACCTTGAAATCTGGAAACAAACCGACCCCAACACC 146  
142 TTCTACGAGGACCTCATTTCTCATGTGCTGAGAGCTGTGATGATGTTTGAAGCTTC 201  
147 TTCTACGAGGACCTCATTTCTCATGTGCTGAGAGCTGTGATGATGTTTGAAGCTTC 206  
202 CTGGAGGTGCTATGAGGAGCAATCCAGAGATCCAGTGTCTGCTGGAGACGTTCTTCAACTGAC 261  
207 CTGGAGGTGCTATGAGGAGCAATCCAGAGATCCAGTGTCTGCTGGAGACGTTCTTCAACTGAC 266  
262 CTGTGATGATCCGTTTGTGCTGTGAGAGTGGCTGTGAGGCAATCTGGGGCTTGTGTAACCAAGAC 321  
267 CTGTGATGATCCGTTTGTGCTGTGAGAGTGGCTGTGAGGCAATCTGGGGCTTGTGTAACCAAGAC 326  
322 CAGATGCGCAAGATGTGAAGCAGTTCTATGACCAAGGCGCTTTCAGCAACCTGTGATGAT 381  
327 CAGATGCGCAAGATGTGAAGCAGTTCTATGACCAAGGCGCTTTCAGCAACCTGTGATGAT 386  
382 GATGATGCGCAACATGCGCAAGCTGTGTGAAGACTTTCATGAGACGCTCAACCTGTGT 441  
387 GATGATGCGCAACATGCGCAAGCTGTGTGAAGACTTTCATGAGACGCTCAACCTGTGT 446  
442 GGCTCCAAACGACCTGACCAACACTGACCTACACCAATCTGAGGAAACAGCCTGTGTCCCTCA 501  
447 GGCTCCAAACGACCTGACCAACACTGACCTACACCAATCTGAGGAAACAGCCTGTGTCCCTCA 506  
502 GGCGGCAACATGCTCAACCCCTTATCTGCAGACCAAGATTGTTCATCAGAAAATCGATGAGCTC 561  
507 GGCGGCAACATGCTCAACCCCTTATCTGCAGACCAAGATTGTTCATCAGAAAATCGATGAGCTC 566  
562 TTCTCTGTGGAAAGCTGTACCTCATTTGGAATTGAGCCATTGTGTACTGTCAATTATGATC 621  
567 TTCTCTGTGGAAAGCTGTACCTCATTTGGAATTGAGCCATTGTGTACTGTCAATTATGATC 626  
622 TTTCAGATGATTTCTGACCATGTGTGCTGTGAGCATCCGAAACAGCTCCGCTGTACTCA 681  
627 TTTCAGATGATTTCTGACCATGTGTGCTGTGAGCATCCGAAACAGCTCCGCTGTACTCA 686  
682 G 682  
887 G 687

RESULT 5  
B1157227  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

B1157227 698 bp mRNA linear EST 05-JUN-2001  
602992435F1 NCI\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5063432 5',  
mRNA sequence.  
B1157227  
B1157227.1 GI:14617228  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Bukatyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 698)  
NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.

FEATURES	Location/Qualifiers
source	1. .698

```

/organism="Mus musculus"
/mol_type="rRNA"
/strain="129, C57BL/6J, FVB/N"
/db_xref="taxon:10090"
/clone_image="5063432"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/clone_idb="NCI_CGAP_Mam3"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."

```

BASE COUNT	149 a	180 c	190 g	179 e
ORIGIN				

ORIGIN

Query Match	96.8%	Score 660	DB 12	Length 698
Best Local Similarity	100.0%	Pred. No. 2.6e-164		
Matches 660; Conservative	0	Mismatches	0	Indels 0; Gaps 0;

QY	23	ATTTCGCTCTTGGCTGAGCGGTATCTTAAGGTGAGCTCTGTGTGGTGTATG	82
Db	1	ATTTCGCTCTTGGCTGAGCGGTATCTTAAGGTGAGCTCTGTGTGGTGTATG	60
QY	83	ATCCACGACACACAGCTGCTGTACCTGGAACTGGGAAACAAACCGGCAACCCAAACCT	142
Db	61	ATCCACGACACACAGCTGCTGTACTGTGAACCTGGGAAACAAACCGGCAACCCAAACCT	120
QY	143	TCTACGTGGGCATCTACATTTCTCATTTGCTGTGGAGCTGTGATGATGTTTGTAGGCTTCC	202
Db	121	TCTACGTGGGCATCTACATTTCTCATTTGCTGTGGAGCTGTGATGATGTTTGTAGGCTTCC	180
QY	203	TGGGGGTCTATGGGGGCATTCACAGAGTCCAGAGTCTGCTGTGGGAGCGTTCTTCACTGGCC	262
Db	181	TGGGGGTCTATGGGGGCATTCACAGAGTCCAGAGTCTGCTGTGGGAGCGTTCTTCACTGGCC	240
QY	263	TTCGTATCTCTGTTTGCTGTGAGGTGGCTGCACAGCATCTGGGGCTTGTGTAACCAAGACC	322
Db	241	TTCGTATCTCTGTTTGCTGTGAGGTGGCTGCACAGCATCTGGGGCTTGTGTAACCAAGACC	300
QY	323	AGATCGCACAGATGTGAAAGCACTTCTATGACCAAGGCCCTTTCACAGACCTGTGTATGATG	382

|||||  
301 AGATCCGCAAGATGTAAGCAGTTTATATACAGGCCCTTCAGCAAGCTGTATGATG 360  
383 ATGATGCGCAATATGCGAAGGCTGTGTGAAGACTTTCATGAGAGCGCTCAACTGTG 442  
361 ATGATGCGCAATATGCGAAGGCTGTGTGAAGACTTTCATGAGAGCGCTCAACTGTG 420  
443 GCTCCAAAGCACTGACCAACATGACTTACCAAGCACTGAGGAAACAGCTGTGCTCAG 502  
421 GCTCCAAAGCACTGACCAACATGACTTACCAAGCACTGAGGAAACAGCTGTGCTCAG 480  
503 GCGGCAACATCTACACCCCTTATCTGAGAGAGATTGTATCAGAAATGATGAGCTCT 562  
481 GCGGCAACATCTACACCCCTTATCTGAGAGAGATTGTATCAGAAATGATGAGCTCT 540  
563 TCTCTGGGAAGCTGTACCTGATTTGGAATTGCAAGCACTGTGTAGCTGTATGATCT 622  
541 TCTCTGGGAAGCTGTACCTGATTTGGAATTGCAAGCACTGTGTAGCTGTATGATCT 600  
623 TTGAGATGATTTGAGCATGTGTGTGCTGTGGGATCCGGAACAGCTCCGTGTACTGAG 682  
601 TTGAGATGATTTGAGCATGTGTGTGCTGTGGGATCCGGAACAGCTCCGTGTACTGAG 660  
B1653596 777 bp mRNA linear EST 12-SEP-2001  
603300230P1 NCI\_CGAP\_Mam3 Mus musculus cDNA clone IMAGE:5340823 5'  
mRNA sequence.  
B1653596  
B1653596.1 GI:15567832  
EST.  
Mus musculus (house mouse)  
Mus musculus  
Bukaryofa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 777)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM1865 row: 1 column: 08  
High quality sequence stop: 760.  
Location/Qualifiers  
1. 777  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="129\_C57BL/6J\_PVB/N"  
/db\_xref="taxon:10090"  
/clone="IMAGE:5340823"  
/tissue\_type="tumor, gross tissue"  
/dev\_stage="10 months"  
/lab\_host="DH10B"  
/clone\_1lb="NCI\_CGAP\_Mam3"  
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;  
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dr.  
Library constructed by Life Technologies. Investigators  
providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH  
Reference for transgenic model: Xu et al., Nature Genetics  
22, 37-43 (1999)."  
COUNT 168 a 200 c 206 g 203 t  
N  
TY Match 96.4%; Score 657.4; DB 12; Length 777;  
t Local Similarity 99.6%; Pred. No. 1.3e-163;

Matches 660; Conservative 0; Mismatches 1; Indels 2; Gaps 2;  
QY 1 AAATACCTGCTCTGCTCTTCAATTGCTTCTTCTGCTGCTGAGGGGTGATCTAGGT 60  
Db 10 AAATACCTGCTCTGCTCTTCAATTGCTTCTTCTGCTGCTGAGGGGTGATCTAGGT 69  
QY 61 GTAGCTGTGTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 120  
Db 70 GTAGCTGTGTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 129  
QY 121 AACAAACCGGACCCCAACACCTTCTAGCTGAGCATCTAATTCTGATGCTGAGAGCT 180  
Db 130 AACAAACCGGATACCCCAACACCTTCTAGCTGAGCATCTAATTCTGATGCTGAGAGCT 189  
QY 181 GTAGATGATGTTGATGAGCTTCTGAGGGGTGCTATGAGGAGCATTCAGAGTCCAGTGTCTG 240  
Db 190 GTAGATGATG-TTGATGAGCTTCTGAGGGGTGCTATGAGGAGCATTCAGAGTCCAGTGTCTG 248  
QY 241 CTGGGAGGCTTCTTCACTGCTGCTTGTGATCTGTTTGTGCTGTGAGGTGAGGCTGAGGCAATC 300  
Db 249 CTGGGAGGCTTCTTCACTGCTGCTTGTGATCTGTTTGTGCTGTGAGGTGAGGCAATC 308  
QY 301 TGGGCTTCTGTAACAAAGACCAAGATGCGCAAGATGGAAGAGTTCTATGACCAAGGCC 360  
Db 309 TGGGCTTCTGTAACAAAGACCAAGATGCGCAAGATGGAAGAGTTCTATGACCAAGGCC 368  
QY 361 CTTCAGCAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 420  
Db 369 CTTCAGCAAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 428  
QY 421 CATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
Db 429 CATGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 488  
QY 481 AGGAACAGCTGTGTCTCTGAGGGGCAAGATGATGATGATGATGATGATGATGATGATG 540  
Db 489 AGGAACAGCTGTGTCTCTGAGGGGCAAGATGATGATGATGATGATGATGATGATGATG 548  
QY 541 CATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600  
Db 549 CATGAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 608  
QY 601 GTGTAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 659  
Db 609 GTGTAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 668  
QY 660 CCGGAACAGCTCCGTGTACTGAG 682  
Db 669 CCGGAACAGCTCCGTGTACTGAG 691  
RESULT 7  
BO714590 956 bp mRNA linear EST 16-JUL-2002  
LOCUS BO714590  
DEFINITION AGENCOURT 8292060 NIH\_MGC\_129 Mus musculus cDNA clone IMAGE:6308902  
ACCESSION BO714590  
VERSION BO714590.1 GI:21853489  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Bukaryofa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 956)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Susan L. Sullivan, PhD.  
cDNA Library Preparation: ResGen, Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation



LT 9	8	852 bp	mRNA	linear	EST 02-APR-2003
8103					
	CBS58103				
	AGNCOCOURT_11381828				
	IMAGE:30244420	5'	Mus	musculus	cdna clone
	CSSTON				
	CS58103				
					mRNA sequence.

JRBS source	Location/Qualifiers
	1. .852

```

    94.68; Score 645; DB 14; Length 852;

```

Oy	1	AAATACCTGCTCTGCTCTTCAATTTGCTCTTGTGCTGAGGCGTATCCTAGT	60
Db	160	AAATACCTGCTCTTGGCTTCAATTTGCTCTTGTGCTGAGGCGTATCCTAGT	215
Oy	61	GTAGCTCTGTGCTGTGCTCATGATCCAGACCAACAGCTGTGTACCTGAACTGGGA	120
Db	220	GTAGCTCTGTGCTGTGCTCATGATCCAGACCAACAGCTGTGTGTACCTGAACTGGGA	275
Oy	121	AAACAAACGGACCCCAACACCTTTCTAATGAGGGACATCTAATTTCTATGCTGTGGAGCT	180
Db	280	AAACAAACGGACCCCAACACCTTTCTAATGAGGGACATCTAATTTCTATGCTGTGGAGCT	335
Oy	181	GTGATGATGTTTGTAGGCTTCTGTGGGTGTATGTGGGCAATCCAGAGTCCAGATGTCTG	240
Db	340	GTGATGATGTTTGTAGGCTTCTGTGGGTGTATGTGGGCAATCCAGAGTCCAGATGTCTG	395
Oy	241	CTGGGGACCTTTCTTCACTGTCTTGTATCTCTTTTCTGTGAGTGTCTGCAGGCATC	300
Db	400	CTGGGGACCTTTCTTCACTGTCTTGTATCTCTTTTCTGTGAGTGTCTGCAGGCATC	455
Oy	301	TGGGGCTTCTGTAAACAAAGACAGATGTGCAGAGATGTGAAGAGCTTCTATGACCAAGCC	350
Db	460	TGGGGCTTGTGTAAACAAAGACAGATGTGCAGAGATGTGAAGAGCTTCTATGACCAAGCC	515
Oy	361	CTTCAAGCAAGCTGTGATGTGATGTATGTATGTCCAAATGTCCAAAGGCTGTGTGAAGATTTCT	420
Db	520	CTTCAAGCAAGCTGTGATGTGATGTATGTATGTCCAAATGTCCAAAGGCTGTGTGAAGATTTCT	575
Oy	421	CATGAGACGCTCAACGTGTTGTGCTTCAACGCCATGACGCCATCTGATACCAACATATCTG	480
Db	580	CATGAGACGCTCAACGTGTTGTGCTTCAACGCCATGACGCCATCTGATACCAACATATCTG	635
Oy	481	AGGAACAGGCTGTGTGCTTCAAGGCGGCAACATATCTCAACCCCTTACTGTGACGACAAATTTGT	540
Db	640	AGGAACAGGCTGTGTGCTTCAAGGCGGCAACATATCTCAACCCCTTACTGTGACGACAAATTTGT	695
Oy	541	CATCAGAAATGATGAGAGCTCTTCTCTGTGGAAAGCTGTATCTCATTTGGAATTTGACGCATTT	600
Db	700	CATCAGAAATGATGAGAGCTCTTCTCTGTGGAAAGCTGTATCTCATTTGGAATTTGACGCATTT	755
Oy	601	GTGTGAGCTGTCAATATGATCTTTGAGATGTATCTTGACGATGTGATCTGTGTGGCATC	660
Db	760	GTGTGAGCTGTCAATATGATCTTTGAGAAAGATTTCTACCAAGGTGTGTGTCTGTCCGAC	815
Oy	661	C 661	
Db	820	C 820	

RESULT	10
BII13178	
LOCUS	
DEFINITION	BII13178 759 bp mRNA linear EST 26-JUN-2001
ACCESSION	M62900106F1 NCI CGAP MamS Mus musculus cDNA clone IMAGE:5030136 5'
VERSION	BII13178
KEYWORDS	BII13178.1 GI:14564079
SOURCE	EST.
ORGANISM	Mus musculus (house mouse)
	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	NIH-MGC http://mgc.nci.nih.gov/.
AUTHORS	1 (base 1 to 759)
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished
COMMENT	Contact: Robert Strauberg, Ph.D. Email: cgabs-r@mail.nih.gov Tissue Procurement: Lohar Hemighausen Ph.D., Robin Humphreys cDNA Library Preparation: Life Technologies, Inc. cDNA library Arrayed by: The I.M.A.G.E. Consortium (LNLN)

DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM1084 row: k column: 01  
 High quality sequence stop: 741.

RES  
 source  
 Location/Qualifiers

1. 759  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5030136"  
 /tissue\_type="tumor, gross tissue"  
 /dev\_stage="7 months"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Mam2"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1; Salt;  
 Site 2: Not; Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies. Investigators  
 providing samples: Lothar Hennighausen/Robin Humphreys,  
 NIH"

COUNT 166 a 194 c 209 g 190 t

ry Match 94.0%; Score 641.2; DB 12; Length 759;  
 Local Similarity 99.1%; Pred. No. 2.6e-159;  
 ches 676; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

1 AATATCTGCTCTTCTGCTCAATTTGCTCTTCTGCTGCTGAGGCTGATCTAGT 60  
 13 AATATCTGCTCTTCTGCTCAATTTGCTCTTCTGCTGCTGAGGCTGATCTAGT 72  
 61 GTAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 73 GTAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 132  
 121 AACAAACCGGACCCCAACACCTTCTAGCTGCTGCTGCTGCTGCTGCTGCTGCT 180  
 133 AACAAACCGGACCCCAACACCTTCTAGCTGCTGCTGCTGCTGCTGCTGCTGCT 192  
 181 GTATGATGATTTGATGAGCTTCTGAGGCTGCTATGAGGCTGCTGAGAGTCCAGTCTG 240  
 193 GTATGATGATTTGATGAGCTTCTGAGGCTGCTATGAGGCTGCTGAGAGTCCAGTCTG 252  
 241 CTGAGGAGCTTCTGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGCT 300  
 253 CTGAGGAGCTTCTGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGCT 312  
 301 TGAGGCTTCTGTAACAAGACAGATGCGCAAGATGTAAGAGCTTCTATGACCAAGGCT 360  
 313 TGAGGCTTCTGTAACAAGACAGATGCGCAAGATGTAAGAGCTTCTATGACCAAGGCT 372  
 361 CTTCAGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420  
 373 CTTCAGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 432  
 421 CATGAGAGCTGCACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 433 CATGAGAGCTGCACTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 492  
 481 AGAAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 493 AGAAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 552  
 541 CATGAGAAATGATGAGCTCTTCTCTGAGAGCTGATCTGATGATGATGATGATGATGAT 600  
 553 CATGAGAAATGATGAGCTCTTCTCTGAGAGCTGATCTGATGATGATGATGATGATGAT 610  
 601 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 611 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 669

Qy 661 CGAAGAGCTCCGTACTAG 682  
 Db 670 CGAAGAGCTCCGTACTAG 691

RESULT 11

LOCUS B1695388 759 bp. mRNA linear EST 18-SEP-2001  
 DEFINITION 603345472P1 NCI\_CGAP\_Mam2 Mus musculus cDNA clone IMAGE:5373184 5',  
 mRNA sequence.  
 ACCESSION B1695388  
 VERSION B1695388  
 KEYWORDS B1695388.1 GI:15658017  
 SOURCE EST.  
 ORGANISM Mus musculus (house mouse)

REFERENCE  
 AUTHORS Mammalia; Butcheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NIH-MGC <http://imgc.ncl.nih.gov/>  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM11949 row: p column: 17  
 High quality sequence stop: 754.

FEATURES

source

1. 759  
 Location/Qualifiers

/organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="FVB/N-3"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:5373184"  
 /tissue\_type="tumor, biopsy sample"  
 /dev\_stage="5 months"  
 /lab\_host="DH10B"  
 /clone\_lib="NCI CGAP Mam2"  
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site: 1; Salt;  
 Site 2: Not; Cloned unidirectionally. Primer: Oligo dt.  
 Library constructed by Life Technologies. Investigator  
 providing samples: Gilbert Smith, NIH"

BASE COUNT 164 a 194 c 204 g 197 t

Query Match 93.8%; Score 640; DB 12; Length 759;  
 Best Local Similarity 99.7%; Pred. No. 5.4e-159;

Matches 662; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 18 CTTCAATTCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 77  
 Db 1 CTTCAATTCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60  
 Qy 78 TCATGATTCACAGACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 137  
 Db 61 TCATGATTCACAGACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120  
 Qy 138 CACCTTCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 197  
 Db 121 CACCTTCTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 178  
 Qy 198 CTTCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 257  
 Db 179 CTTCTGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 238  
 Qy 258 CTGCTTGTGATCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 317  
 Db 239 CTGCTTGTGATCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 298



318 AGACGATGCGGAGGATGAGAGGAGTTCTATGACGAGGCCCTTCAGCAAGCTGTAT 377  
 |||||  
 299 AGACGATGCGGAGGATGAGAGGAGTTCTATGAGCAGGAGGCCCTTCAGCAAGCTGTAT 358  
 |||||  
 378 GAGATGATGATGCGCAACATGCGCAAGGCTGTGTGAGAGACTTTTCATGAGAGCTCAACTG 437  
 |||||  
 359 GAGATGATGATGCGCAACATGCGCAAGGCTGTGTGAGAGACTTTTCATGAGAGCTCAACTG 418  
 |||||  
 438 TTGTGCTCCAAAGCACTGACCACTGACTGACCAACCATATCTAGAGAAAGCTGTGTCC 497  
 |||||  
 419 TTGTGCTCCAAAGCACTGACCACTGACTGACCAACCATATCTAGAGAAAGCTGTGTCC 478  
 |||||  
 498 CTGAGGCGGCAACATGCTGACCCCTTACTGCGAGGAATTTGATCAGAAAATGATGA 557  
 |||||  
 479 CTGAGGCGGCAACATGCTGACCCCTTACTGCGAGGAATTTGATCAGAAAATGATGA 538  
 |||||  
 558 GCTCTTCTGCGAAGCTGACCTGATGATGGAATTGAGCACTGTGTGTGATGAT 617  
 |||||  
 539 GCTCTTCTGCGAAGCTGATGATGGAATTGAGCACTGTGTGTGATGAT 598  
 |||||  
 618 GATCTTTGATGATGATTTCTGAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 677  
 |||||  
 599 GATCTTTGATGATGATTTCTGAGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGT 658  
 |||||  
 678 CTGA 681  
 |||||  
 659 CTGA 662

LT 12  
 6993  
 S BO886993 903 bp mRNA linear EST 16-AUG-2002  
 NITTON AGENCOURT\_8752226 NIH\_MGC\_130 Mus musculus cDNA clone IMAGE:633531  
 5', mRNA sequence.  
 SSION BO886993  
 ION BO886993.1 GI:22279007  
 ORDS EST.  
 CR Mus musculus (house mouse)  
 GANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 903)  
 NIH-MGC http://mgs.nci.nih.gov/  
 TLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 UNPUBLISHED  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Mark Maconochie, Ph.D. and Nancy L. Freeman,  
 Ph.D.  
 cDNA Library Preparation: ResGen, Invitrogen Corp  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.llnl.gov  
 Plate: LLM13792 row: 0 column: 04  
 High quality sequence stop: 644.  
 Location/Qualifiers  
 1..903

ORES  
 source  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:10090"  
 /clone="IMAGE:633531"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_1ib="NIH MGC 130"  
 /note="Organ: oocytes; Vector: pCMV-SPORT6.1.ccd; Site 1: EcoRV, Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.95 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH MGC library."  
 COUNT 181 a 265 c 237 g 215 t 5 others  
 IN

Query Match 93.8%; Score 639.4; DB 13; Length 903;  
 Best Local Similarity 99.5%; Pred. No. 8.5e-159;  
 Matches 640; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AAATACCTGCTCTTCTGCTTCAATTTGCTTCTGCTGCTGAGGAGCTGATCTTGA 60  
 |||||  
 Db 236 AAATACCTGCTCTTCTGCTTCAATTTGCTTCTGCTGCTGAGGAGCTGATCTTGA 295  
 |||||  
 QY 61 GTAGCTGT 120  
 |||||  
 Db 296 GTAGCTGT 355  
 |||||  
 QY 121 AACAAACCGGACCCCAACACCTTTAGCTGAGGATCTAATTTGATGCTGTGAGAGCT 180  
 |||||  
 Db 356 AACAAACCGGACCCCAACACCTTTAGCTGAGGATCTAATTTGATGCTGTGAGAGCT 415  
 |||||  
 QY 181 GTATGATGATTTGT 240  
 |||||  
 Db 416 GTATGATGATTTGT 475  
 |||||  
 QY 241 CTGGGAGAGCTTCTTCAACCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300  
 |||||  
 Db 476 CTGGGAGAGCTTCTTCAACCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 535  
 |||||  
 QY 301 TGGGAGCTTGTAAACAAAGACAGATGCGCAAGATGTGAAGCACTTCAATGACAGGCC 360  
 |||||  
 Db 536 TGGGAGCTTGTAAACAAAGACAGATGCGCAAGATGTGAAGCACTTCAATGACAGGCC 595  
 |||||  
 QY 361 CTTGAGCAAGCTGT 420  
 |||||  
 Db 536 CTTGAGCAAGCTGT 655  
 |||||  
 QY 421 CATGAGAGCTGCTCACTGT 480  
 |||||  
 Db 656 CATGAGAGCTGCTCACTGT 715  
 |||||  
 QY 481 AGGAACAGCTGT 540  
 |||||  
 Db 716 AGGAACAGCTGT 775  
 |||||  
 QY 541 CATGAGAAATGATGAGCTTCTTCTGAGAACTGTACTGATGTGAATGACCACTT 600  
 |||||  
 Db 776 CATGAGAAATGATGAGCTTCTTCTGAGAACTGTACTGATGTGAATGACCACTT 835  
 |||||  
 QY 601 GTGTAGCTGTATATGATGATCTTTGAGATGATCTGAGACATG 643  
 |||||  
 Db 836 GTGTAGCTGTATATGATGATCTTTGAGATGATCTGAGACATG 878  
 |||||

RESULT 13  
 BF533191 642 bp mRNA linear EST 11-DEC-2000  
 LOCUS 602073837/71 NCI\_GCAP\_L19 Mus musculus cDNA clone IMAGE:4211024 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BF533191  
 VERSION BF533191.1 GI:1162054  
 KEYWORDS EST.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 642)  
 NIH-MGC http://mgs.nci.nih.gov/  
 TLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 UNPUBLISHED  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue Procurement: Jeffrey E. Green, M.D.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be



	495	GACGACACTGCATACCAACATATAGAGGA CAGCCTGTGCCCTCAGGGCGAACTACT	515
	436	GACGACACTGCATACCAACATATAGAGGA CAGCCTGTGCCCTCAGGGCGAACTACT	495
	516	CACCCCTTACTGACGAGAATTGTTCATGAAAATGATGAGCTCTTCTCTGGAGGCT	575
	496	CACCCCTTACTGACGAGAATTGTTCATGAAAATGATGAGCTCTTCTCTGGAGGCT	555
	576	GTACCTCATTTGGAATTGAGGCATTGTGTGTGCTGTGATTATGATCTTTGAGATGTTCT	635
	556	GTACCTCATTTGGAATTGAGGCATTGTGTGTGCTGTGATTATGATCTTTGAGATGTTCT	615
	636	GAGCATGTGCTGTCGCTGTGTCATTCGGAAACAGCTCCGTACTGAG	682
	616	GAGCATGTGCTGTCGCTGTGTCATTCGGAAACAGCTCCGTACTGAG	662
LRT 15			
/S	1041		
NITION	601805089P1 NCI_GCAP_Mam5 Mus musculus cDNA clone IMAGE:4035767 5'	806 bp mRNA linear EST 31-OCT-2000	
	mRNA sequence.		
SSION	BF181041		
ION	BF181041		
ORDS	BF181041.1 GI:11059183		
	EST.		
CE	Mus musculus (house mouse)		
GANISM	Mus musculus		
RENCE	Bukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;		
TIR	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
THORS	1 (bases 1 to 806)		
URNAL	NIH-MGC http://mgc.nci.nih.gov/		
ENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished		
	Contact: Robert Straubeberg, Ph.D.		
	Email: cgabds@remail.nih.gov		
	Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys		
	cDNA Library Preparation: Life Technologies, Inc.		
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLT)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/ILNLT at:		
	http://image.llnl.gov		
	plate: ILAM9310 row: b column: 24		
	High quality sequence stop: 742.		
URES			
SOURCE	1. 806		
	Location/Qualifiers		

Db	74	GTAGCTCTGTGGTGTGGTCAATGACAGACCAACAGCTGTGTACCTGTGAACCTGGAA	13
Qy	121	AAAGAAACCGGACACCCAAACACCTTCTTAAGTGGGCATCTACATTTCTATGTCTGTGGAGCT	180
Db	134	AAAGAAACCGGACCCAAACACCTTCTTAAGTGGGCATCTTAATTTCTCA--TGCTGTGGAGCT	192
Qy	181	GTGATGATGTTTGTAGAGCTTCCGAGGGGTGCTATGGGGGCATCCAGAGAGTCCAGTGTG	240
Db	193	GTGATGATGTTTGTAGAGCTTCCGAGGGGTGCTATGGGGGCATCCAGAGAGTCCAGTGTGCTG	252
Qy	241	CTGAGGACGTTCTTCACTGCTGTGTGATTCCTGTTTGGCTGTGAGGTGGCTGCAGGCATC	300
Db	253	CTGAGGACGTTCTTCACTGCTGTGTGATTCCTGTTTGGCTGTGAGGTGGCTGCAGGCATC	312
Qy	301	TGGGGCTGTGTAAAGAAAGACAGATCCGCAAGATGTGAAGCAAGTTCTATGACAGGCC	360
Db	313	TGGGGCTGTGTAAAGAAAGACAGATCCGCAAGATGTGAAGCAAGTTCTATGACAGGCC	372
Qy	361	CTTCAGCAAGCTGTGATGTGATGATATGCAATATGCAAGGCTGTGTGTGAAGCACTTTC	420
Db	373	CTTCAGCAAGCTGTGATGTGATGATATGCAATATGCAAGGCTGTGTGTGAAGCACTTTC	432
Qy	421	CATGAGACGCTCAACTGTTTGTGGCTTCCACGCACTGACCAACTGACTTACCAACCATCTG	480
Db	433	CATGAGACGCTCAACTGTTTGTGGCTTCCACGCACTGACCAACTGACTTACCAACCATCTG	492
Qy	481	AGGAAAGAGCTGTGTCCCGAGGGGGCAACATATCAACCCCTTAATGTGACGAAGATGT	540
Db	493	AGGAAAGAGCTGTGTCCCGAGGGGGCAACATATCAACCCCTTAATGTGACGAAGATGT	551
Qy	541	CATCAGAAAATGATGAGTCTTCTCTGTGGAGAGCTGTAACCTCATTTGATGTGACGCAAT	600
Db	552	CATCAGAAAATGATGAGTCTTCTCTGTGGAGAGCTGTAACCTCATTTGATGTGACGCAAT	609
Qy	601	GTGTGTAGCTGTCAATTATGATCTTTGAGATGATTTCTGACATGTGTGTGTGTGTGATC	660
Db	610	GTGTGTAGCTGTCAATTATGATCTTTGAGATGATTTCTGACATGTGTGTGTGTGTGATC	669
Qy	661	CGGAAAGCTCCGTTACTGAG	682
Db	670	CGGAAAG--TCCGTGTACTGAG	690

Search completed: December 9, 2003, 08:29:05  
Job time : 2630 secs

COUNT	176 a	186 c	224 g	220 t
217 Match		92.0%;	Score 627.2;	DB 10;
31 Local Similarity		98.8%;	Pred. No. 1.4e-155;	
674; Conservative		0;	Mismatches 3;	Indels 5; Gaps 4
1	AAATACCTGCTCTTCGCTTCATTTGCTTTCTGAGCTGAGAGCGTATCCTAGGT	60		
14	AAATACCTGCTCTTCGCTTCATTTGCTTTCTGAGCTGAGAGCGTATCCTAGGT	73		
61	GTAGCTGTGTGCTGTGATTCACAGACACGAGCTGCTGTAACCTGGAACCTGGA	120		

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

protein - protein search, using sw model

December 9, 2003, 08:29:17 ; Search time 74 Seconds

(without alignments)  
484.760 Million cell updates/sec

US-10-035-914-2

1 KILLFVNFVFWIAGGVILG.....IFEMILSMVLCCGINSSVY 226

Gapop 10.0 , Gapext 0.5

1 number of hits satisfying chosen parameters: 1107863

Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description
1177	99.7	258	24	ABG76341
1103	93.5	236	22	Recombinant mouse
1103	93.5	236	23	Chronic hepatitis
1103	93.5	236	23	Human CD81 protein
1103	93.5	236	24	Human tetraspanin
1103	93.5	236	24	Human expressed pr
1103	93.5	236	24	Human expressed pr
1103	93.5	236	24	Human expressed pr
1082	91.7	236	20	Human CD81m protei

10	1082	91.7	236	24	ABU07307
11	1034	87.6	262	21	AAB53254
12	1034	87.6	262	21	ABU07304
13	1034	87.6	279	22	AA689276
14	1034	87.6	279	22	ABU07306
15	1003	85.0	222	24	ABG76342
16	1003	85.0	222	24	ABG76344
17	518.5	43.9	227	17	AAB66834
18	518.5	43.9	227	17	AAB76366
19	518.5	43.9	227	23	AAB14636
20	518.5	43.9	227	24	ABU05057
21	518.5	43.9	227	24	ABU05060
22	518.5	43.9	228	22	ABBA4581
23	518.5	43.9	228	24	ABU05048
24	518.5	43.9	228	24	ABU05050
25	518.5	43.9	228	24	ABU05052
26	518.5	43.9	228	24	ABU05053
27	518.5	43.9	228	24	ABU05059
28	518.5	43.9	228	24	ABU05059
29	518.5	43.9	275	21	AAB43936
30	518.5	43.9	275	22	AA675156
31	518.5	43.9	275	24	ABU05054
32	513.5	43.5	228	13	AA675525
33	513.5	43.5	228	24	ABU05056
34	512.5	43.4	228	24	ABU05049
35	510.5	43.3	226	23	ABBA4580
36	510.5	43.3	226	23	ABBA4580
37	453.5	38.4	221	19	AA61623
38	453.5	38.4	221	24	ABU59143
39	453.5	38.4	229	22	AA673745
40	425	36.0	454	22	ABG14067
41	424.5	36.0	221	24	ABU69142
42	403	34.2	113	20	AA09155
43	403	34.2	113	24	ABU07309
44	401	34.0	116	20	AA09154
45	401	34.0	116	24	ABU07308

ALIGNMENTS

RESULT 1  
ID ABG76341 standard; Protein: 258 AA.  
AC ABG76341:  
DT 10-MAY-2003 (first entry)  
XX  
XX  
DE Recombinant mouse protein, CD81.  
XX  
XX Mouse; protein targeting; exosome; lactadherin; C1 domain; C2 domain;  
KW membrane vesicle; mutant; mulein.  
XX  
OS Mus sp.  
OS Synthetic.  
OS  
PN MO2003016522-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002MO-EP09108.  
XX  
PR 17-AUG-2001; 2001US-113159P.  
PR 26-DEC-2001; 2001US-343991P.  
PA (ANOS-) ANOSYS INC.  
XX  
XX Delcayre A, Le Pecq J;  
PI WPI; 2003-268331/26.  
DR  
XX  
XX Targeting polypeptides to exosomes providing a chimeric genetic

construct and introducing the construct into exosome-producing cells in vivo or ex vivo

Example 6; Page 87-88; 94pp; English.

The present invention relates to a method and compounds for targeting polypeptides to exosomes. The method comprises providing a chimeric polypeptide construct encoding the polypeptide fused to a targeting polypeptide comprising lactadherin or its portion comprising a functional C1 and/or C2 domain, and introducing the construct into exosome-producing cells in vivo or ex vivo, to generate recombinant exosomes. The method is useful for targeting proteins to membrane vesicles, particularly exosomes, and is useful in experimental, research, therapeutic, prophylactic, and diagnostic areas. The present sequence represents a recombinant mouse protein.

Sequence 258 AA;

CY Match 99.7%; Score 1177; DB 24; Length 258;  
Local Similarity 99.6%; Pred. No. 5e-127; Indels 0; Gaps 0;  
Seqs 223; Conservative 1; Mismatches 0;

```

1 KYLLFVFNFMVFLAGVILGVALLMRHPDPTSLVLELGNKPAPNTFYGIYILAVGA 60
  |||||
11 KYLLFVFNFMVFLAGVILGVALLMRHPDPTSLVLELGNKPAPNTFYGIYILAVGA 70
  |||||
61 VMFVFGFLGCGAIOESQCLGTFTCLVILFACVAAAGVFNKQIADVKQFYDQA 120
  |||||
71 VMFVFGFLGCGAIOESQCLGTFTCLVILFACVAAAGVFNKQIADVKQFYDQA 130
  |||||
121 LQAAVMDDDANNKAVKTFHETLNCSSNALTTTLTILNSLCPSGNLTPLAQDC 180
  |||||
131 LQAAVMDDDANNKAVKTFHETLNCSSNALTTTLTILNSLCPSGNLTPLAQDC 190
  |||||
181 HOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGIRNSSVY 226
  |||||
191 HOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGIRNSSVY 236
  |||||

```

1 2  
826  
AAG64826 standard; protein; 236 AA.

AAG64826;

20-SEP-2001 (first entry)

Chronic hepatitis treatment related protein SEQ ID NO: 7.

Chronic hepatitis; viral antigenic protein; hepatitis C; hepatitis B.

Homo sapiens.

MO200147545-A1.

05-JUL-2001.

28-DEC-2000; 2000MO-JP09393.

28-DEC-1999; 99JP-0374087.

(SUMU) SUMITOMO PHARM CO LTD.

Tohdoh N, Murata M, Enjoji T;

WPI; 2001-425585/45.

Treatment and prevention of chronic hepatitis

Example 1; Page 73-74; 128pp; Japanese.

The present invention describes a method of preventing and treating chronic hepatitis, involving administering an oligopeptide which

CC (a) has binding affinity towards the viral antigenic protein;  
CC (b) inhibits binding affinity of the virus towards the receptor protein  
CC of the target cell and  
CC (c) has analogy with the receptor protein at the amino acid level. This  
CC can be used to prevent and treat hepatitis B and C. The present sequence  
CC is a protein described in the exemplification of the invention.  
XX  
SQ Sequence 236 AA;

Query Match 93.5%; Score 1103; DB 22; Length 236;  
Best Local Similarity 91.6%; Pred. No. 1.5e-118;  
Matches 207; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 KYLLFVFNFMVFLAGVILGVALLMRHPDPTSLVLELGNKPAPNTFYGIYILAVGA 60
  |||||
DB 11 KYLLFVFNFMVFLAGVILGVALLMRHPDPTSLVLELGNKPAPNTFYGIYILAVGA 70
  |||||
QY 61 VMFVFGFLGCGAIOESQCLGTFTCLVILFACVAAAGVFNKQIADVKQFYDQA 120
  |||||
DB 71 VMFVFGFLGCGAIOESQCLGTFTCLVILFACVAAAGVFNKQIADVKQFYDQA 130
  |||||
QY 121 LQAAVMDDDANNKAVKTFHETLNCSSNALTTTLTILNSLCPSGNLTPLAQDC 180
  |||||
DB 131 LQAAVMDDDANNKAVKTFHETLNCSSNALTTTLTILNSLCPSGNLTPLAQDC 190
  |||||
QY 181 HOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGIRNSSVY 226
  |||||
DB 191 HOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGIRNSSVY 236
  |||||

```

RESULT 3

ABP52845  
ID ABP52845 standard; Protein; 236 AA.

XX ABP52845;

DT 04-NOV-2002 (first entry)

XX Human CD81 protein SEQ ID NO:5.

XX Human; p15 region; chromosome 11; tumour growth; infertility; cytostatic;  
XX hepatitis C virus infection; anti-infectivity; gene therapy; CD 81;  
XX cluster of differentiation antigen 81.

XX Homo sapiens.

XX MO200261085-A2.

XX 08-AUG-2002.

XX 31-OCT-2001; 2001MO-US45381.

XX 31-OCT-2000; 2000US-244705P.

XX (RYAN/) RYAN J W.

XX Ryan JW;

XX WPI; 2002-619251/66.

XX N-PSDB; ABQ75121.

XX New genes obtainable from the p15 region of human chromosome 11 (e.g.  
XX human achaete-scute homolog 2), useful in gene therapy, particularly  
XX for preventing or treating tumor growth, infertility or hepatitis C  
XX virus infection -

XX Claim 1; Fig 1; 94pp; English.

XX The present invention describes an isolated genomic polynucleotide (I),  
XX which is obtainable from the p15 region of human chromosome 11. (HSH2),  
XX Specifically described are the human achaete-scute homologue 2 (HSH2),  
XX SMS3, tumour suppressing sub-transferable candidate 6 (TSSC6), ribosomal  
XX protein L26 (RIBO26), cluster of differentiation antigen 81 (CD 81) and

tumour suppressing sub-transferable candidate 4 (TSSC4) genes as given in ABQ75117 to ABQ75122, encoding the proteins given in ABP52841 to ABP52846. Also described is an isolated polynucleotide obtainable from the p15 region of chromosome 11 having the sequence, which comprises any of two nucleotide sequences given in ABQ75123 and ABQ75124. (1) has cytostatic and anti-fertility activities and can be used in gene therapy. The HNSH2, SMS3, TSSC6, RIBO26, CD 81 or TSSC4 polynucleotide sequences can be used for preventing, treating or ameliorating a medical condition, or for manufacturing a gene therapy for the prevention, treatment or amelioration of such medical conditions. An antisense oligonucleotide or mimetic of (1) can also be useful for manufacturing a medicament for the prevention, treatment or amelioration of these medical conditions. These conditions include tumour growth, infertility, or hepatitis C virus infection.

Sequence 236 AA:

ery Match 93.5%; Score 1103; DB 23; Length 236;  
at Local Similarity 91.6%; Pred. No. 1.5e-118;  
tches 207; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

```

1 KYLLFVFNFWMLAGGVILGVALLMRHDPQTSLYLTLGNKPAPNTFYVGIYILAVGA 60
  |||||
11 KYLLFVFNFWMLAGGVILGVALLMRHDPQTSLYLTLGNKPAPNTFYVGIYILAVGA 70
  |||||
61 VMFVFGFGCGAIGSOCLGTFPTCLVILFACVAAAGVGFVNKOIADVKQFYDQA 120
  |||||
71 VMFVFGFGCGAIGSOCLGTFPTCLVILFACVAAAGVGFVNKOIADVKQFYDQA 130
  |||||
121 LQQAAMDDANNAAVVKTFHETLNCSSNALTTTLTLRNSLCPSSGNILTPLLQDDC 180
  |||||
131 LQQAAMDDANNAAVVKTFHETLNCSSSTLTALTSTVAKNNLCPSSGNISLTLFKEDC 190
  |||||
181 HOKIDELFSKGLYLIGIAIVAVVIMIFEMILSMVLCGGIRNSSY 226
  |||||
191 HOKIDELFSKGLYLIGIAIVAVVIMIFEMILSMVLCGGIRNSSY 236
  |||||

```

XT 4

1499 AAG78499 standard; protein; 236 AA.

AAG78499;

19-JUN-2002 (first entry)

Human tetraspanin CD81.

Human; CD81; tetraspanin; protein coordinate data.

Homo sapiens.

Key Location/Qualifiers  
Domain 113..201  
/note="extracellular loop"

MO200202632-A2.

10-JAN-2002.

03-JUL-2001; 2001WO-1B01458.

03-JUL-2000; 2000GB-0016362.

(CHIR-) CHIRON SPA.

Bolognesi M, Grandi G;

WPI; 2002-154731/20.

Modulating (M1) the ability of a tetraspanin protein to form a dimer of the wild-type TTS, by mutating at least one of the amino acids corresponding to specific human CD81 residues

XX Disclosure; Page 2; 56pp; English.

PS This invention relates to the modulation of the ability of a tetraspanin  
XX (TTS) protein to form a dimer with the wild-type TTS, comprising  
XX mutating at least one of the amino acids corresponding to human CD81  
XX residues, 114, 119, 123, 125, 126, 129, 142, 146, 149, 150, 153, 154,  
XX 197, 198, 199 and 200. Methods used are useful for modulating the  
XX ability of a TTS protein to form a dimer with the wild-type TTS and  
XX are useful for identifying a compound that inhibits or promotes TTS  
XX dimerization. A computer-based method is also used for identifying a  
XX ligand which can interact with TTS to promote or inhibit its ability to  
XX dimerise. Certain products from the methods are useful as a  
XX pharmaceutical, for treating a patient. This sequence represents  
XX human CD81, a tetraspanin.

Sequence 236 AA:

Query Match 93.5%; Score 1103; DB 23; Length 236;  
Best Local Similarity 91.6%; Pred. No. 1.5e-118;  
Matches 207; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 KYLLFVFNFWMLAGGVILGVALLMRHDPQTSLYLTLGNKPAPNTFYVGIYILAVGA 60
  |||||
DB 11 KYLLFVFNFWMLAGGVILGVALLMRHDPQTSLYLTLGNKPAPNTFYVGIYILAVGA 70
  |||||
QY 61 VMFVFGFGCGAIGSOCLGTFPTCLVILFACVAAAGVGFVNKOIADVKQFYDQA 120
  |||||
DB 71 VMFVFGFGCGAIGSOCLGTFPTCLVILFACVAAAGVGFVNKOIADVKQFYDQA 130
  |||||
QY 121 LQQAAMDDANNAAVVKTFHETLNCSSNALTTTLTLRNSLCPSSGNILTPLLQDDC 180
  |||||
DB 131 LQQAAMDDANNAAVVKTFHETLNCSSSTLTALTSTVAKNNLCPSSGNISLTLFKEDC 190
  |||||
QY 181 HOKIDELFSKGLYLIGIAIVAVVIMIFEMILSMVLCGGIRNSSY 226
  |||||
DB 191 HOKIDELFSKGLYLIGIAIVAVVIMIFEMILSMVLCGGIRNSSY 236
  |||||

```

RESULT 5

ID ABU07302 standard; Protein; 236 AA.

XX AC ABU07302;

DT 29-JAN-2003 (first entry)

XX Human expressed protein tag (EPT) #2003.

XX Translational profiling; expressed protein tag; EPT; kinase;

KW phosphatase; protease; protease inhibitor; transporter;

KW cytoskeletal protein; receptor; transcription factor; cancer; MHC;

KW major histocompatibility complex; myeloma; colon cancer;

XX gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;

XX leukemia.

XX Homo sapiens.

XX OS MO200202632-A2.

XX PD 10-OCT-2002.

XX PF 28-MAR-2002; 2002WO-US09671.

XX PR 28-MAR-2001; 2001US-279495P.

XX PR 21-MAY-2001; 2001US-292544P.

XX PR 08-AUG-2001; 2001US-310801P.

XX PR 01-OCT-2001; 2001US-326370P.

XX PR 04-DEC-2001; 2001US-336780P.

XX PR 20-FEB-2002; 2002US-358985P.

XX PA (ZYCO-) ZYCO INC.

Chicz RM, Tomlinson AJ, Urban RG;

WPI; 2003-040607/03.

New polypeptides (e.g. kinases, phosphatases, proteases, transporters, cytoskeletal proteins, receptors or transcription factors), useful for treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or leukemia

Example 2; SEQ ID No 2003; 134pp; English.

The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor. The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that binds to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukaemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling.

Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 236 AA;

Query Match 93.5%; Score 1103; DB 24; Length 236;  
Local Similarity 91.6%; Pred. No. 1.5e-118;

Shes 207; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

```

1 KTLFVFNFMVFLAGVYILGVALLMLRHPDPTSLLYLRLGNKPAFNTFYVGIYLLIIVGA 60
11 KTLFVFNFMVFLAGVYILGVALLMLRHPDPTSLLYLRLGNKPAFNTFYVGIYLLIIVGA 70
61 VMFVGFGLCYGAIQESQCLGFTFTCLVILFACVVAAGIWMGFVNQDIADVKQFYDQA 120
71 VMFVGFGLCYGAIQESQCLGFTFTCLVILFACVVAAGIWMGFVNQDIADVKQFYDQA 130
121 LQQAIVDDDDANNAKAAVKTFTHTLNCSSNALTTTLTILRNSLCPSGGNITLPLLODC 180
131 LQQAIVDDDDANNAKAAVKTFTHTLNCSSNALTTTLTILRNSLCPSGGNITLPLLODC 190
181 HOKIDELFSGKLYLIGIAIYVAIVIMIFEMILSMVLCGGRNSSVY 226
191 HOKIDDLFSGKLYLIGIAIYVAIVIMIFEMILSMVLCGGRNSSVY 236

```

103 ABU07303 standard; Protein; 236 AA.

ABU07303;

19-JAN-2003 (first entry)

human expressed protein tag (EPT) #2004.

Translational profiling; expressed protein tag; EPT; kinase; phosphatase; protease; protease inhibitor; transporter; cytoskeletal protein; receptor; transcription factor; cancer; MHC; major histocompatibility complex; myeloma; colon cancer; gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma; leukemia.

homo sapiens.

MO200278524-A2.

```

XX 10-OCT-2002.
PD 28-MAR-2002; 2002MO-US09671.
XX 28-MAR-2001; 2001US-279495P.
PR 21-MAY-2001; 2001US-292544P.
XX 08-AUG-2001; 2001US-310801P.
PR 01-OCT-2001; 2001US-326370P.
XX 04-DEC-2001; 2001US-336780P.
PR 20-FEB-2002; 2002US-358985P.
XX (ZYCO-) ZYCO5 INC.
XX Chicz RM, Tomlinson AJ, Urban RG;
PI WPI; 2003-040607/03.
XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters,
PT cytoskeletal proteins, receptors or transcription factors), useful for
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma
PT or leukemia
XX Example 2; SEQ ID No 2004; 134pp; English.
XX The invention describes a purified polypeptide, which comprises a
CC fragment of a kinase, phosphatase, protease, protease inhibitor,
CC transporter, cytoskeletal protein, receptor or transcription factor.
CC The polypeptide is useful as an immunogenic composition for eliciting
CC in a mammal an immunogenic response directed against any of the purified
CC polypeptide. The purified polypeptide, or the antibody that binds to
CC this polypeptide, is useful for treating cancer. The polypeptide is
CC also useful for identifying compounds that binds to a naturally
CC processed class I or class II MHC-binding polypeptide. The polypeptides
CC and polynucleotides are particularly useful for treating or preventing
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,
CC lymphoma or leukaemia. These are also useful for screening agents for
CC treating the above mentioned diseases. This sequence represents an
CC expressed protein tag (EPT) isolated from human tissue for translational
CC profiling.
CC Note: This sequence does not appear in the printed specification but was
CC obtained in electronic format directly from WIPO at
CC ftp://wipo.int/pub/published\_pct\_sequences.
XX
XX Sequence 236 AA;
SQ
Query Match 93.5%; Score 1103; DB 24; Length 236;
Best Local Similarity 91.6%; Pred. No. 1.5e-118;
Matches 207; Conservative 14; Mismatches 5; Indels 0; Gaps 0;
QY 1 KTLFVFNFMVFLAGVYILGVALLMLRHPDPTSLLYLRLGNKPAFNTFYVGIYLLIIVGA 60
DB 11 KTLFVFNFMVFLAGVYILGVALLMLRHPDPTSLLYLRLGNKPAFNTFYVGIYLLIIVGA 70
QY 61 VMFVGFGLCYGAIQESQCLGFTFTCLVILFACVVAAGIWMGFVNQDIADVKQFYDQA 120
DB 71 VMFVGFGLCYGAIQESQCLGFTFTCLVILFACVVAAGIWMGFVNQDIADVKQFYDQA 130
QY 121 LQQAIVDDDDANNAKAAVKTFTHTLNCSSNALTTTLTILRNSLCPSGGNITLPLLODC 180
DB 131 LQQAIVDDDDANNAKAAVKTFTHTLNCSSNALTTTLTILRNSLCPSGGNITLPLLODC 190
QY 181 HOKIDELFSGKLYLIGIAIYVAIVIMIFEMILSMVLCGGRNSSVY 226
DB 191 HOKIDDLFSGKLYLIGIAIYVAIVIMIFEMILSMVLCGGRNSSVY 236

```

RESULT 7

ABU07305

ID ABU07305 standard; Protein; 236 AA.

XX ABU07305;

XX

29-JAN-2003 (first entry)

Human expressed protein tag (EPT) #2006.

Translational profiling; expressed protein tag; EPT; kinase; phosphatase; protease; protease inhibitor; transporter; cytoskeletal protein; receptor; transcription factor; cancer; MHC; major histocompatibility complex; myeloma; colon cancer; gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma; leukemia.

Homo sapiens.

WO200278524-A2.

10-OCT-2002.

28-MAR-2002; 2002WO-US09671.

28-MAR-2001; 2001US-279495P.

21-MAY-2001; 2001US-292544P.

08-AUG-2001; 2001US-310801P.

01-OCT-2001; 2001US-326370P.

04-DEC-2001; 2001US-336780P.

20-FEB-2002; 2002US-358985P.

(ZYCO-) ZYCOS INC.

Chicz RM, Tomlinson AJ, Urban RG;

WPI; 2003-040607/03.

New polypeptides (e.g. kinases, phosphatases, proteases, transporters, cytoskeletal proteins, receptors or transcription factors), useful for treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or leukemia.

Example 2; SEQ ID No 2006; 134pp; English.

The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor. The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that binds to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling.

Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 236 AA;

try Match 93.5%; Score 1103; DB 24; Length 236;

at Local Similarity 91.6%; Pred. No. 1,56-118;

ches 207; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

```

1 KTLFVFNFEVLGAGVILGVALMLRHPDPTSLYLEGNKPAENTFYVGIYIIAAG 60
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
11 KTLFVFNFEVLGAGVILGVALMLRHPDPTSLYLEGNKPAENTFYVGIYIIAAG 70
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 VMFVGFPGCGAIGESQCLIGTFTCLVILPACVAAIGWGFVNKDQIADVQKPYDQA 120
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
71 VMFVGFPGCGAIGESQCLIGTFTCLVILPACVAAIGWGFVNKDQIADVQKPYDQA 130
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 LQQAAMDDANNAAKAVVTFHETLNCGSSNMLTTLTTLILNSLCPSSGNILTPILQDPC 180

```

```

Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
131 LQQAAMDDANNAAKAVVTFHETLNCGSSSTLTALTSLVLRNNLCPSSGNILSNLFEKDC 190

```

```

QY |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 HOKIDELPSGKLYLGIAIYVAVMIFEMILSMWLCCGINSSSY 226
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
191 HOKIDELPSGKLYLGIAIYVAVMIFEMILSMWLCCGINSSSY 236

```

RESULT 8

ABU07310 standard; Protein; 236 AA.

ABU07310;

29-JAN-2003 (first entry)

Human expressed protein tag (EPT) #2011.

Translational profiling; expressed protein tag; EPT; kinase;

phosphatase; protease; protease inhibitor; transporter;

cytoskeletal protein; receptor; transcription factor; cancer; MHC;

major histocompatibility complex; myeloma; colon cancer;

gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;

leukemia.

Homo sapiens.

WO200278524-A2.

10-OCT-2002.

28-MAR-2002; 2002WO-US09671.

28-MAR-2001; 2001US-279495P.

21-MAY-2001; 2001US-292544P.

08-AUG-2001; 2001US-310801P.

01-OCT-2001; 2001US-326370P.

04-DEC-2001; 2001US-336780P.

20-FEB-2002; 2002US-358985P.

(ZYCO-) ZYCOS INC.

Chicz RM, Tomlinson AJ, Urban RG;

WPI; 2003-040607/03.

New polypeptides (e.g. kinases, phosphatases, proteases, transporters, cytoskeletal proteins, receptors or transcription factors), useful for treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or leukemia.

Example 2; SEQ ID No 2011; 134pp; English.

The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor. The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that binds to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling.

Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at [ftp://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 236 AA;



ry Match 93.5%; Score 1103; DB 24; Length 236;  
 Local Similarity 91.6%; Pred. No. 1.5e-118;  
 ches 207; Conservative 14; Mismatches 5; Indels 0; Gaps 0;  
 1 KYLLFVNFVFMVLGGVILGVALLMLRHDPTTSLYLTELGNKPAFPTFYGIYLLAVGA 60  
 11 KYLLFVNFVFMVLGGVILGVALLMLRHDPTTSLYLTELGNKPAFPTFYGIYLLAVGA 70  
 61 VMFVFGFLGCTGAIQESQCLLTGFTCLVILFACFVAAGIWMGFVNKOIADKVKQFYDOA 120  
 71 VMFVFGFLGCTGAIQESQCLLTGFTCLVILFACFVAAGIWMGFVNKOIADKVKQFYDOA 130  
 121 LQAAVMDDDANNAKAVVKTFFHTLNCSSNALTTTLTILNSLCPSGNIIPTLLQDDC 180  
 131 LQAAVMDDDANNAKAVVKTFFHTLNCSSNALTTTLTILNSLCPSGNIIPTLLQDDC 190  
 181 HOKIDELFSGKLYLIGIAIVAVVIMFEMILSMVLCCGIRNSSV 226  
 191 HOKIDELFSGKLYLIGIAIVAVVIMFEMILSMVLCCGIRNSSV 236

153  
 AAY09153 standard; Protein; 236 AA.  
 AAY09153;

07-JUL-1999 (first entry)

human CD81m protein sequence.

CD81 protein; receptor; Hepatitis C virus; HCV; E2 protein; infection; immunogen; therapy; diagnosis; EC2 region.

homo sapiens.

MO9918198-A1.

15-APR-1999.

06-OCT-1998; 98MO-IB01628.

03-JUN-1998; 98GB-0013560.

06-OCT-1997; 97GB-0021182.

(CHIR-) CHIRON SPA.

brignani S, Grandi G;

PI; 1999-264018/22.

Hepatitis C receptor protein CD81

claim 2; Fig 1; 55pp; English.

The invention relates to the use of the CD81 protein that is identified as a receptor for Hepatitis C virus (HCV). Compounds that specifically bind to CD81 can be used for therapy and diagnosis of HCV. The CD81 protein binds to HCV, preferably to the E2 protein. The CD81 protein can be used to raise antibodies, which can then be used in the diagnosis or therapy of HCV. The antibodies can be used to prevent the virus binding to patient cells and being internalized. The CD81 protein can also be used to treat HCV infection. The CD81 protein, a functional equivalent, or a compound that binds to CD81, can be used in the manufacture of a medicament for the treatment and diagnosis of a HCV infection. The CD81 protein can also be used as a protective immunogen in the control of HCV. The present sequence represents the human CD81 protein (SWISSPROT Accn No: P18582; EMBL/GENBANK Accn No: M33680).  
 Note: The EMBL/GENBANK Accn number is indicated wrongly as M33680 in the claims; the correct accession number is M33680 for the human CD81 protein.

SQ Sequence 236 AA;  
 Query Match 91.7%; Score 1082; DB 20; Length 236;  
 Best Local Similarity 90.7%; Pred. No. 4e-116;  
 Matches 204; Conservative 14; Mismatches 7; Indels 0; Gaps 0;  
 1 KYLLFVNFVFMVLGGVILGVALLMLRHDPTTSLYLTELGNKPAFPTFYGIYLLAVGA 60  
 11 KYLLFVNFVFMVLGGVILGVALLMLRHDPTTSLYLTELGNKPAFPTFYGIYLLAVGA 70  
 61 VMFVFGFLGCTGAIQESQCLLTGFTCLVILFACFVAAGIWMGFVNKOIADKVKQFYDOA 120  
 71 VMFVFGFLGCTGAIQESQCLLTGFTCLVILFACFVAAGIWMGFVNKOIADKVKQFYDOA 130  
 121 LQAAVMDDDANNAKAVVKTFFHTLNCSSNALTTTLTILNSLCPSGNIIPTLLQDDC 180  
 131 LQAAVMDDDANNAKAVVKTFFHTLNCSSNALTTTLTILNSLCPSGNIIPTLLQDDC 190  
 181 HOKIDELFSGKLYLIGIAIVAVVIMFEMILSMVLCCGIRNSSV 225  
 191 HOKIDELFSGKLYLIGIAIVAVVIMFEMILSMVLCCGIRNSSV 235

RESULT 10  
 ABU07307  
 ID ABU07307 standard; Protein; 236 AA.  
 AC ABU07307;  
 XX  
 XX  
 XX  
 DT 29-JAN-2003 (first entry)  
 XX  
 DE Human expressed protein tag (EPT) #2008.

XX  
 XX Translational profiling; expressed protein tag; EPT; kinase;  
 KW phosphatase; protease; protease inhibitor; transporter;  
 KW cytoskeletal protein; receptor; transcription factor; cancer; MEC;  
 KW major histocompatibility complex; myeloma; colon cancer;  
 KW gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;  
 KW leukemia.

XX Homo sapiens.

XX MO200278524-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US09671.

XX 28-MAR-2001; 2001US-279495P.

XX 21-MAY-2001; 2001US-292544P.

XX 08-AUG-2001; 2001US-310801P.

XX 01-OCT-2001; 2001US-326370P.

XX 04-DEC-2001; 2001US-336780P.

XX 20-FEB-2002; 2002US-358985P.

XX (ZYCO-) ZYCO INC.

XX Chicx RM, Tomlinson AJ, Urban RG;

XX WPI; 2003-040607/03.

XX New polypeptides (e.g. kinases, phosphatases, proteases, transporters, cytoskeletal proteins, receptors or transcription factors), useful for treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or leukemia

XX Example 2: SEQ ID No 2008; 134pp; English.

The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor. The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified

polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that bind to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling.

Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 236 AA;

Query Match 91.7%; Score 1082; DB 24; Length 236;  
Local Similarity 90.7%; Pred. No. 4e-116;  
Conservative 14; Mismatches 7; Indels 0; Gaps 0;

```

1 KTLFVFNFMVLAGVILGVALMLRHDQTSILYLKGNRPAPNTFYVGIYILAVGA 60
11 KTLFVFNFMVLAGVILGVALMLRHDQTSILYLKGNRPAPNTFYVGIYILAVGA 70
61 VMFVGLGCTGAGIAGSQCCLGTFCTVILFACVAGAGVFNKQIADVKQFYDQA 120
71 VMFVGLGCTGAGIAGSQCCLGTFCTVILFACVAGAGVFNKQIADVKQFYDQA 130
121 LQAAVDDDDANNAKAVVTFHETLNCSSNALTTTLTILNSLCPSSGNITLPLAQDC 180
131 LQAAVDDDDANNAKAVVTFHETLNCSSNALTTTLTILNSLCPSSGNITLPLAQDC 190
181 HOKIDELFSGKLYLIGIAIIVAVIMIFEMILSMVLCGIRNSV 225
191 HOKIDELFSGKLYLIGIAIIVAVIMIFEMILSMVLCGIRNSV 235
AAB53254 standard; Protein; 262 AA.

```

09-MAR-2001 (first entry)

Human colon cancer antigen protein sequence SEQ ID NO:794.

Human; colon cancer; colon cancer antigen; diagnosis; detection; identification; cytostatic; cardioactive; neuroprotective; vulnary; immunomodulatory; muscular; gynaecological; gastrointestinal; nephrotropic; antiinfective; antibacterial; gene therapy; wound; neural disorder; immune system disorder; muscular disorder; reproductive disorder; gastrointestinal disorder; renal disorder; infectious disease; cardiovascular disorder.

Homo sapiens.

W0200055351-A1.

21-SEP-2000.

08-MAR-2000; 2000WO-US05883.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI; 2000-587534/55.

N-PSDB; AAC98011.

Colon cancer associated gene sequences, referred to as colon cancer

PT antigens, useful for the treatment, prevention, and diagnosis of colon disorders such as colon cancer -

PS Claim 11; Page 1350-1351; 21049P; English.

AAC97991 to AAC98763 encode the human colon cancer associated proteins, called human colon cancer antigens, given in AAB53234 to AAB54006. The human colon cancer antigens can have cytostatic, cardioactive, muscular, neuroprotective, immunomodulatory, gynaecological, gastrointestinal, vulnary, nephrotropic, antiinfective and antibacterial activities, and can be used in gene therapy. The colon cancer antigen polynucleotides, proteins and antibodies to the proteins are useful for the prevention, treatment and diagnosis of colon disorders, such as colon cancer. The polynucleotides may be used in diagnostics and research, such as for chromosome identification, and as hybridisation probes. The proteins may also be used to prevent diseases such as neural disorders, immune system disorders, muscular disorders, reproductive disorders, gastrointestinal disorders, wounds, renal disorders, infectious diseases, and cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent sequences used in the exemplification of the present invention.

Sequence 262 AA;

Query Match 87.6%; Score 1034; DB 21; Length 262;  
Local Similarity 91.1%; Pred. No. 1.6e-110;  
Conservative 14; Mismatches 5; Indels 0; Gaps 0;

```

13 LAGVILGVALMLRHDQTSILYLKGNRPAPNTFYVGIYILAVGAVMFMVGLGCTG 72
49 LAGVILGVALMLRHDQTSILYLKGNRPAPNTFYVGIYILAVGAVMFMVGLGCTG 108
73 AIGSQCCLGTFCTVILFACVAGAGVFNKQIADVKQFYDQAQAAVDDDDANN 132
109 AIGSQCCLGTFCTVILFACVAGAGVFNKQIADVKQFYDQAQAAVDDDDANN 168
133 AKAVVTFHETLNCSSNALTTTLTILNSLCPSSGNITLPLAQDC HOKIDELFSGK 192
169 AKAVVTFHETLNCSSNALTTTLTILNSLCPSSGNITLPLAQDC HOKIDELFSGK 228
193 YLIGIAIIVAVIMIFEMILSMVLCGIRNSV 226
229 YLIGIAIIVAVIMIFEMILSMVLCGIRNSV 262

```

RESULT 12

ID AABU07304 standard; Protein; 262 AA.

AC AABU07304;

DT 29-JAN-2003 (first entry)

DB Human expressed protein tag (EPT) #2005.

XX Translation profiling; expressed protein tag; EPT; kinase;  
XX phosphatase; protease; protease inhibitor; transporter;  
XX cytoskeletal protein; receptor; transcription factor; cancer; MHC;  
XX major histocompatibility complex; myeloma; colon cancer;  
XX gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma;  
XX leukemia.

OS Homo sapiens.

XX W0200278524-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-US09671.

XX 28-MAR-2001; 2001US-279495P.

XX 21-MAY-2001; 2001US-292544P.

PR 08-AUG-2001; 2001US-310801P.

01-OCT-2001; 2001US-326370P.  
04-DEC-2001; 2001US-336780P.  
20-FEB-2002; 2002US-358985P.  
(ZYCO-) ZYCOs INC.  
Chicz RM, Tomlinson AJ, Urban RG;  
WPI; 2003-040607/03.

New polypeptides (e.g. kinases, phosphatases, proteases, transporters, cytoskeletal proteins, receptors or transcription factors), useful for treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or leukemia

Example 2; SEQ ID No 2005; 134pp; English.

The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor. The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that binds to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling.  
Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 262 AA;

Match 87.6%; Score 1034; DB 24; Length 262;  
Local Similarity 91.1%; Pred. No. 1.6e-110;  
Matches 195; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

```
13 LAGGVLGVALLMLRHPDPTSLLYELGNKPAPNTFYGIYLLIIVGAVMMFVGLGCG 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
49 LAGGVLGVALLMLRHPDPTSLLYELGNKPAPNTFYGIYLLIIVGAVMMFVGLGCG 108
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
73 AIOESQCLLGTFTCLVILPACVAAAGIWFVFNKQDIADVKQFYDQALQAAVDDANN 132
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
109 AIOESQCLLGTFTCLVILPACVAAAGIWFVFNKQDIADVKQFYDQALQAAVDDANN 168
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
133 AKAVVKTFTHTLCCGSNALTTTLTILRNSLCPSSGNITLPLLODCHOKIDLFSGKL 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
169 AKAVVKTFTHTLCCGSSTLTALTTSVLRNNLCPSSGNITLPLFKEDCHOKIDLFSGKL 228
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
193 YLIGIAIIVAVVIMIFEMILSMVLCGGIRNSSY 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
229 YLIGIAIIVAVVIMIFEMILSMVLCGGIRNSSY 262
```

F 13  
276  
YAG89276 standard; Protein; 279 AA.  
YAG89276;

11-SRP-2001 (first entry)

human secreted protein, SEQ ID NO: 396.

human; secreted protein; gene therapy; vaccine; treatment; diagnosis;  
JENSET.  
homo sapiens.

```
XX
PN MO200142451-A2.
XX
XX 14-JUN-2001.
PD
XX
XX 07-DEC-2000; 2000MO-IB01938.
PF
XX
XX 08-DEC-1999; 99US-0169629.
PR
XX 06-MAR-2000; 2000US-0187470.
XX
XX (GENT ) GENSET.
PA
PI Dumas Milne Edwards J, Bougueleret L, Jobert S;
PI WPI; 2001-367870/38.
XX
XX N-PSDB; AAH64879.
DR
```

Full length GENSET human nucleic acid encoding potentially secreted proteins, useful in gene therapy and vaccination against a variety of diseases, and for diagnosis of those diseases -  
Claim 21; Page 881; 921pp; English.

The invention relates to full length GENSET human nucleic acids encoding potentially secreted proteins. The nucleic acids and the polypeptides of they encode may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate GENSET gene expression. For example, they be used to treat disorders associated with decreased GENSET gene expression by rectifying mutations or deletions in a patient's genome that affect the activity of GENSET or by supplementing the patient's own production of GENSET polypeptides. Conversely, CC antiense nucleic acid molecules may be administered to down regulate GENSET expression by binding with the cells' own genes and preventing their expression. The sense and antisense nucleic acids may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acid sequences in samples, and hence to determine which patients may be in need of restorative therapy. CC The GENSET polypeptides may be used as antigens in the production of antibodies and in assays to identify modulators (agonists and CC antagonists) of GENSET polypeptide expression and activity. The CC present sequence is a GENSET polypeptide of the invention.

Sequence 279 AA;

Query Match 87.6%; Score 1034; DB 22; Length 279;  
Best Local Similarity 91.1%; Pred. No. 1.7e-110;  
Matches 195; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

```
QY 13 LAGGVLGVALLMLRHPDPTSLLYELGNKPAPNTFYGIYLLIIVGAVMMFVGLGCG 72
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 66 LAGGVLGVALLMLRHPDPTSLLYELGNKPAPNTFYGIYLLIIVGAVMMFVGLGCG 125
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 73 AIOESQCLLGTFTCLVILPACVAAAGIWFVFNKQDIADVKQFYDQALQAAVDDANN 132
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 126 AIOESQCLLGTFTCLVILPACVAAAGIWFVFNKQDIADVKQFYDQALQAAVDDANN 185
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 133 AKAVVKTFTHTLCCGSNALTTTLTILRNSLCPSSGNITLPLLODCHOKIDLFSGKL 192
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 186 AKAVVKTFTHTLCCGSSTLTALTTSVLRNNLCPSSGNITLPLFKEDCHOKIDLFSGKL 245
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 193 YLIGIAIIVAVVIMIFEMILSMVLCGGIRNSSY 226
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 246 YLIGIAIIVAVVIMIFEMILSMVLCGGIRNSSY 279
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

RESULT 14

ABU07306  
ID ABU07306 standard; Protein; 279 AA.  
XX  
XX ABU07306;  
AC  
XX  
XX 29-JAN-2003 (first entry)  
DT  
XX

Human expressed protein tag (EPT) #2007.

Translational profiling; expressed protein tag; EPT; kinase; phosphatase; protease; protease inhibitor; transporter; cytoskeletal protein; receptor; transcription factor; cancer; MHC; major histocompatibility complex; myeloma; colon cancer; gastric cancer; adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

Homo sapiens.

WO200278524-A2.

10-OCT-2002.

28-MAR-2002; 2002WO-US09671.

28-MAR-2001; 2001US-279495P.

21-MAY-2001; 2001US-292544P.

08-AUG-2001; 2001US-310801P.

01-OCT-2001; 2001US-326370P.

04-DEC-2001; 2001US-336780P.

20-FEB-2002; 2002US-358985P.

(ZYCO-) ZYCOS INC.

Chicz RM, Tomlinson AJ, Urban RG;

WPI; 2003-040607/03.

Example 2; SEQ ID No 2007; 134pp; English.

The invention describes a purified polypeptide, which comprises a fragment of a kinase, phosphatase, protease, protease inhibitor, transporter, cytoskeletal protein, receptor or transcription factor. The polypeptide is useful as an immunogenic composition for eliciting in a mammal an immunogenic response directed against any of the purified polypeptide. The purified polypeptide, or the antibody that binds to this polypeptide, is useful for treating cancer. The polypeptide is also useful for identifying compounds that binds to a naturally processed class I or class II MHC-binding polypeptide. The polypeptides and polynucleotides are particularly useful for treating or preventing myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma, lymphoma or leukemia. These are also useful for screening agents for treating the above mentioned diseases. This sequence represents an expressed protein tag (EPT) isolated from human tissue for translational profiling. Note: This sequence does not appear in the printed specification but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 279 AA;

Query Match 87.6%; Score 1034; DB 24; Length 279; Best Local Similarity 91.1%; Pred. No. 1.7e-110; Mismatches 14; Indels 0; Gaps 0; Conserved 195; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

13 LAGGVILGVALMLRHDPTTSLVLELGNKRAPRTFYVGIYILAVGA VMVFGLCTG 72  
66 LAGGVILGVALMLRHDPTTSLVLELGNKRAPRTFYVGIYILAVGA VMVFGLCTG 125  
73 AIOESOCILGFTFCLVILFACVVAAGIWFVNQDAKQVDFDQALQOAVDDANN 132  
126 AIOESOCILGFTFCLVILFACVVAAGIWFVNQDAKQVDFDQALQOAVDDANN 185  
133 AKAVVKTFFHETLNCSSNALTTTLTILRNSLCPSGGINLTPLLQODCHQKIDELFSGKL 192  
186 AKAVVKTFFHETLNCSSNALTTTLTILRNSLCPSGGINLTPLLQODCHQKIDELFSGKL 245

QY 193 YIIGIAIYVAIVIMIFEMILSMVLCGGIRNSSVY 226  
DB 246 YIIGIAIYVAIVIMIFEMILSMVLCGGIRNSSVY 279

RESULT 15

ABG76342  
ID ABG76342 standard; Protein; 222 AA.

AC ABG76342;

DT 10-MAY-2003 (first entry)

DE Truncated form of recombinant mouse protein, CD81.

KW Mouse; protein targeting; exosome; lactadherin; C1 domain; C2 domain; membrane vesicle; mutant; mutein.

OS Mus sp.

PN WO2003016522-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-EP09108.

PR 17-AUG-2001; 2001US-313159P.

PR 26-DEC-2001; 2001US-343991P.

PA (ANOS-) ANOSYS INC.

PI Delcayre A, Le Pecq J;

DR WPI; 2003-268331/26.

PT Targeting polypeptides to exosomes providing a chimeric genetic construct and introducing the construct into exosome-producing cells in vivo or ex vivo

PS Example 6; Page 89; 94pp; English.

CC The present invention relates to a method and compounds for targeting polypeptides to exosomes. The method comprises providing a chimeric genetic construct encoding the polypeptide fused to a targeting polypeptide comprising lactadherin or its portion comprising a functional C1 and/or C2 domain, and introducing the construct into exosome-producing cells in vivo or ex vivo, to generate recombinant vesicles. The method is useful for targeting proteins to membrane vesicles, particularly exosomes, and is useful in experimental research, therapeutic, prophylactic, and diagnostic areas. The present sequence represents the truncated form of a recombinant mouse protein.

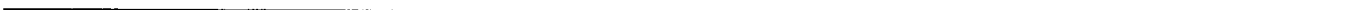
Sequence 222 AA;

Query Match 85.0%; Score 1003; DB 24; Length 222; Best Local Similarity 99.5%; Pred. No. 4.7e-107; Mismatches 189; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KYLLFVNFVFWLAGGVILGVALMLRHDPTTSLVLELGNKRAPRTFYVGIYILAVGA 60  
DB 11 KYLLFVNFVFWLAGGVILGVALMLRHDPTTSLVLELGNKRAPRTFYVGIYILAVGA 70  
QY 61 VMVFVGLCTGAIIOESOCILGFTFCLVILFACVVAAGIWFVNQDAKQVDFDQALQOAVDDANN 120  
DB 71 VMVFVGLCTGAIIOESOCILGFTFCLVILFACVVAAGIWFVNQDAKQVDFDQALQOAVDDANN 130  
QY 121 LQOAVDDANNKAVVKTFFHETLNCSSNALTTTLTILRNSLCPSGGINLTPLLQODC 180  
DB 131 LQOAVDDANNKAVVKTFFHETLNCSSNALTTTLTILRNSLCPSGGINLTPLLQODC 190

181 HOKIDELFSG 190  
| | | | | | | |  
191 HOKIDELFSG 200

h completed: December 9, 2003, 10:29:41  
ime : 77 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

protein - protein search, using sw model

on: December 9, 2003, 10:28:23 / Search time 28 Seconds  
(without alignments)  
341.509 Million cell updates/sec

le: US-10-035-914-2

lect score: 1180  
1 KYLLFVFNFMVLAAGVILG.....IFEMILSMVLCGIRNSVY 226

ing table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

ched: 328717 seqs, 4231058 residues

l number of hits satisfying chosen parameters: 328717

num DB seq length: 0  
num DB seq length: 2000000000

-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

base :

1: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/2/iaa/5A\_COMB.pep:\*  
4: /cgn2\_6/prodata/2/iaa/5B\_COMB.pep:\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/2/iaa/backfillseq1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

o.	Score	Query Match	Length	ID	Description
1	518.5	43.9	227	1	US-08-254-493-1 Sequence 1, Appli
2	518.5	43.9	227	2	US-08-253-751-6 Sequence 6, Appli
3	518.5	43.9	227	2	US-08-453-925-6 Sequence 6, Appli
4	518.5	43.9	227	4	US-08-403-253A-6 Sequence 6, Appli
5	518.5	43.9	227	4	US-08-435-816A-6 Sequence 6, Appli
6	518.5	43.9	227	1	US-08-408-222B-1 Sequence 1, Appli
7	298	25.3	237	3	US-08-808-148-3 Sequence 3, Appli
8	297.5	25.2	219	2	US-08-855-140-4 Sequence 4, Appli
9	272.5	23.1	267	3	US-08-430-225A-20 Sequence 20, Appli
10	265.5	22.5	219	2	US-08-855-140-3 Sequence 3, Appli
11	265.5	22.5	219	2	US-08-807-044-3 Sequence 3, Appli
12	265.5	22.5	219	5	PCT-US91-04966-2 Sequence 2, Appli
13	248.5	21.1	253	3	US-09-333-599-4 Sequence 3, Appli
14	248.5	21.1	253	4	US-09-499-781-4 Sequence 4, Appli
15	246.5	20.9	236	3	US-08-705-771-19 Sequence 19, Appli
16	243.5	20.6	253	3	US-09-333-599-2 Sequence 2, Appli
17	243.5	20.6	253	4	US-09-499-781-2 Sequence 1, Appli
18	238.5	20.2	280	2	US-08-855-140-1 Sequence 1, Appli
19	236	20.0	265	2	US-08-807-044-1 Sequence 1, Appli
20	220.5	18.7	241	3	US-08-808-148-1 Sequence 11, Appli
21	220.5	18.7	241	3	US-09-020-956-114 Sequence 11, Appli
22	220.5	18.7	241	3	US-09-030-607-114 Sequence 11, Appli
23	220.5	18.7	241	4	US-09-439-313-114 Sequence 11, Appli
24	220.5	18.7	241	4	US-09-352-616A-114 Sequence 11, Appli
25	220.5	18.7	241	4	US-09-232-149A-114 Sequence 11, Appli
26	191.5	16.2	252	3	US-08-705-771-17 Sequence 4, Appli
27	187.5	15.9	281	3	US-08-808-148-4 Sequence 4, Appli

28	181.5	15.4	245	4	US-09-482-273-133 Sequence 133, App
29	159	13.5	101	3	US-08-905-223-443 Sequence 443, App
30	113.5	9.6	204	4	US-09-149-476-429 Sequence 429, App
31	105	8.9	260	3	US-08-957-130-15 Sequence 15, Appli
32	100	8.5	258	3	US-08-957-130-13 Sequence 13, Appli
33	87	7.4	362	4	US-09-252-991A-27993 Sequence 27993, A
34	86	7.3	79	3	US-08-630-172-8 Sequence 8, Appli
35	86	7.3	79	3	US-09-375-419-8 Sequence 8, Appli
36	83	7.0	333	4	US-09-170-496D-8 Sequence 8, Appli
37	83	7.0	333	4	US-09-170-496D-168 Sequence 168, App
38	83	7.0	403	3	US-09-328-352-7791 Sequence 7791, App
39	77.5	6.6	362	3	US-08-513-974B-374 Sequence 374, App
40	77	6.5	60	3	US-09-188-930-181 Sequence 181, App
41	77	6.5	60	3	US-09-188-930-320 Sequence 320, App
42	77	6.5	60	4	US-09-312-283C-181 Sequence 181, App
43	77	6.5	60	4	US-09-312-283C-320 Sequence 320, App
44	77	6.5	505	4	US-09-328-352-7155 Sequence 7155, App
45	77	6.5	993	1	US-07-977-451-4 Sequence 4, Appli

#### ALIGNMENTS

RESULT 1  
US-08-254-493-1  
Sequence 1, Application US/08254493  
Patent No. 5439886  
GENERAL INFORMATION:  
APPLICANT: IKENAGA, SHUICHI  
APPLICANT: KOYAMA, MASARU  
APPLICANT: MIYAKE, MASAYUKI  
APPLICANT: SANO, MASARU  
TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSER: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS & CUSHMAN  
STREET: 130 WATER STREET  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: US  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/254,493  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/865552  
FILING DATE: 09-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 079996-1991  
FILING DATE: 12-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 085396-1991  
FILING DATE: 17-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 022321-1992  
FILING DATE: 07-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: RESNICK, DAVID S.  
REGISTRATION NUMBER: 34235  
REFERENCE/DOCKET NUMBER: 41777  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 523-3400  
TELEFAX: (617) 523-6440  
TELEX: 200291 STR UR  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:

LENGTH: 227 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULAR TYPE: protein  
 -254-493-1

RY Match 43.9%; Score 518.5; DB 1; Length 227;  
 Local Similarity 44.9%; Pred. No. 6.8e-52;  
 Cons 101; Conservative 39; Mismatches 72; Indels 13; Gaps 5;

1 KYLLFVNFVFWLGGVILGVALMLRHPDQTSLLYLELNKRPAPNTFYGIYILAVGA 60  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 10 KYLLFGNFIFWLAGIYALGMLRFPDSQTKSIFPEQETNN--NSFTYGVYILIGAGA 67  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 61 VMFVFGTGYGAIQESQCLLGTFFCLVILFACVAAIGFVFNKQDIADVKQFYDQA 120  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 68 LMLVGFELGCCGAVQESQCLGLFPGFLVIFALIAALWGSYHDEVIKEVOEFYKDT 127  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 121 LQAVMDDANNAKAVKTFHEFTLNCCGSNALTTLTTLIR--NSLCPSSGNILTPLLQ 178  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 128 YNKKTDEPO--RETLKAHYALNCCG-----LAGGVQFISDICK-KDVLFTTVK 178  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

179 DCHQKIDELFSGKLYLIGAIYAAVAVIMIFEMILSMVLCGIRNS 223  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 179 SCPDAIKEVFENKFIHIGAVGIGIYAVVIMIFGIMFSLICCAIRN 223

T 2  
 -253-751-6  
 nence 6, Application US/08253751  
 ent No. 5858358

# GENERAL INFORMATION:

APPLICANT: June, Carl H.  
 APPLICANT: Thompson, Craig B.  
 APPLICANT: Nabel, Gary J.  
 APPLICANT: Gray, Gary S.  
 APPLICANT: Remmert, Paul D.  
 APPLICANT: Freeman, Gordon J.  
 TITLE OF INVENTION: METHODS FOR SELECTIVELY STIMULATING  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD  
 STREET: 60 STATE STREET, SUITE 510  
 CITY: BOSTON  
 STATE: MA  
 COUNTRY: USA

ZIP: 02109

## COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: ASCII TEXT

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/253,751  
 FILING DATE: 3 JUNE 1994

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/073,223  
 FILING DATE: 4 JUNE 1993  
 APPLICATION NUMBER: US 08/200,947  
 FILING DATE: 23 FEB 1994  
 APPLICATION NUMBER: US 07/864,805  
 FILING DATE: 7 APR 1992  
 APPLICATION NUMBER: US 08/247,505  
 FILING DATE: 23 MAY 1994  
 APPLICATION NUMBER: US 07/864,866  
 FILING DATE: 7 APR 1992  
 APPLICATION NUMBER: 08/218,155  
 FILING DATE: 25 MAR 1994  
 APPLICATION NUMBER: US 07/864,807  
 FILING DATE: 7 APR 1992  
 APPLICATION NUMBER: US 07/902,467

FILING DATE: 16 JUNE 1992  
 APPLICATION NUMBER: US 07/275,433  
 FILING DATE: 23 NOV 1988  
 ATTORNEY/AGENT INFORMATION:  
 NAME: MANDRAGOURAS, AMY E.  
 REGISTRATION NUMBER: 36,207  
 REFERENCE/DOCKET NUMBER: RPI-002CPB  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-7400  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 227 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULAR TYPE: protein  
 US-08-253-751-6

Query Match 43.9%; Score 518.5; DB 2; Length 227;  
 Best Local Similarity 44.9%; Pred. No. 6.8e-52;  
 Matches 101; Conservative 39; Mismatches 72; Indels 13; Gaps 5;

QY 1 KYLLFVNFVFWLGGVILGVALMLRHPDQTSLLYLELNKRPAPNTFYGIYILAVGA 60  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 10 KYLLFGNFIFWLAGIYALGMLRFPDSQTKSIFPEQETNN--NSFTYGVYILIGAGA 67  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 QY 61 VMFVFGTGYGAIQESQCLLGTFFCLVILFACVAAIGFVFNKQDIADVKQFYDQA 120  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 68 LMLVGFELGCCGAVQESQCLGLFPGFLVIFALIAALWGSYHDEVIKEVOEFYKDT 127  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 QY 121 LQAVMDDANNAKAVKTFHEFTLNCCGSNALTTLTTLIR--NSLCPSSGNILTPLLQ 178  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 128 YNKKTDEPO--RETLKAHYALNCCG-----LAGGVQFISDICK-KDVLFTTVK 178  
 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 QY 179 DCHQKIDELFSGKLYLIGAIYAAVAVIMIFEMILSMVLCGIRNS 223  
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 179 SCPDAIKEVFENKFIHIGAVGIGIYAVVIMIFGIMFSLICCAIRN 223

## RESULT 3

US-08-453-925-6  
 Sequence 6, Application US/08453925  
 Patent No. 5883223

## GENERAL INFORMATION:

APPLICANT: June, Carl H., Thompson, Craig B., Nabel, Gary J.  
 APPLICANT: Gray, Gary S., Remmert, Paul D., Freeman, Gordon J.  
 TITLE OF INVENTION: METHODS FOR SELECTIVELY STIMULATING  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:

ADDRESSEE: LAHIVE & COCKFIELD  
 STREET: 60 STATE STREET, SUITE 510  
 CITY: BOSTON  
 STATE: MA  
 COUNTRY: USA

ZIP: 02109

## COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: ASCII TEXT

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/453,925  
 FILING DATE: 30 MAY 1995

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/253,751  
 FILING DATE: 3 JUNE 1994  
 APPLICATION NUMBER: US 08/073,223  
 FILING DATE: 4 JUNE 1993  
 APPLICATION NUMBER: US 08/200,947  
 FILING DATE: 23 FEB 1994  
 APPLICATION NUMBER: US 07/864,805  
 FILING DATE: 7 APR 1992

APPLICATION NUMBER: US 08/247,505  
FILING DATE: 23 MAY 1994  
APPLICATION NUMBER: US 07/864,866  
FILING DATE: 7 APR 1992  
APPLICATION NUMBER: 08/218,155  
FILING DATE: 25 MAR 1994  
APPLICATION NUMBER: US 07/864,807  
FILING DATE: 7 APR 1992  
APPLICATION NUMBER: US 07/902,467  
FILING DATE: 16 JUNE 1992  
APPLICATION NUMBER: US 07/275,433  
FILING DATE: 23 NOV 1988  
ATTORNEY/AGENT INFORMATION:  
NAME: MANDRAGOURAS, AMY B.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: RPI-002CPB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 227 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
1-453-925-6

Query Match 43.9%; Score 518.5; DB 2; Length 227;  
Local Similarity 44.9%; Pred. No. 6.8e-52;  
Conservative 39; Mismatches 72; Indels 13; Gaps 5;  
1 KYLLFVFNFMVLGAGVILGVALMLRHPDPTSLYLLEGNKRAPTEVGYILLAVGA 60  
10 KYLLFGNFIPLMAGLAVLAIGMLRFDSTKSIPEQETNN--NSSFTGVYILLGAGA 67  
61 VMNFVGLGCTGAIQESQCLGTFPTCLVILFACGEVAGIWFVFNKQIAKVQFYDQA 120  
68 LMLLVGLGCGGAVQESQCLGTFPTCLVILFACGEVAGIWFVFNKQIAKVQFYDQA 127  
121 LQAVVMDDDANNAKAVKTFHTLNCSSNALTLTTTLR--NSLCPSGNIITPLDQ 178  
128 YNKLKTKDPPQ--RETLKAHYALNCCG-----LAGVQEPISDIPK-KVLEFTVVK 178  
179 DGHQKIDELPSGKLYLIGLAIIVAVIMIFEMILSMVLCCGIRNS 223  
179 SCPDALKEVFNKFIHIGAVGIGIAVMIFGMIFSMILCCAIRN 223

T 4  
1-403-253A-6  
ence 6, Application US/08403253A  
ent No. 6352694  
GENERAL INFORMATION:  
APPLICANT: June, Carl H., Thompson, Craig B., Nabel, Gary J.  
APPLICANT: Gray, Gary S., Remmert, Paul D.  
TITLE OF INVENTION: Methods For Selectively Stimulating Proliferation Of T-Cells  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSER: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/403,253A  
FILING DATE: March 10, 1995  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/253,964  
FILING DATE: 3 JUNE 1994  
APPLICATION NUMBER: US 08/073,223  
FILING DATE: 4 JUNE 1993  
APPLICATION NUMBER: US 08/200,947  
FILING DATE: 23 FEB 1994  
APPLICATION NUMBER: US 07/864,805  
FILING DATE: 7 APR 1992  
APPLICATION NUMBER: US 08/247,505  
FILING DATE: 23 MAY 1994  
APPLICATION NUMBER: US 07/864,866  
FILING DATE: 7 APR 1992  
APPLICATION NUMBER: US 08/218,155  
FILING DATE: 25 MAR 1994  
APPLICATION NUMBER: US 07/864,807  
FILING DATE: 7 APR 1992  
APPLICATION NUMBER: US 07/902,467  
FILING DATE: 16 JUNE 1992  
APPLICATION NUMBER: US 07/275,433  
FILING DATE: 23 NOV 1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Mandragouras, Amy B.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: RPI-002CP2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 227 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-403-253A-6

Query Match 43.9%; Score 518.5; DB 4; Length 227;  
Best Local Similarity 44.9%; Pred. No. 6.8e-52;  
Matches 101; Conservative 39; Mismatches 72; Indels 13; Gaps 5;  
1 KYLLFVFNFMVLGAGVILGVALMLRHPDPTSLYLLEGNKRAPTEVGYILLAVGA 60  
10 KYLLFGNFIPLMAGLAVLAIGMLRFDSTKSIPEQETNN--NSSFTGVYILLGAGA 67  
61 VMNFVGLGCTGAIQESQCLGTFPTCLVILFACGEVAGIWFVFNKQIAKVQFYDQA 120  
68 LMLLVGLGCGGAVQESQCLGTFPTCLVILFACGEVAGIWFVFNKQIAKVQFYDQA 127  
121 LQAVVMDDDANNAKAVKTFHTLNCSSNALTLTTTLR--NSLCPSGNIITPLDQ 178  
128 YNKLKTKDPPQ--RETLKAHYALNCCG-----LAGVQEPISDIPK-KVLEFTVVK 178  
179 DGHQKIDELPSGKLYLIGLAIIVAVIMIFEMILSMVLCCGIRNS 223  
179 SCPDALKEVFNKFIHIGAVGIGIAVMIFGMIFSMILCCAIRN 223

RESULT 5  
US-08-435-816A-6  
Sequence 6, Application US/08435816A  
Patent No. 6534055  
GENERAL INFORMATION:  
APPLICANT: June, Carl H.  
APPLICANT: Thompson, Craig B.  
APPLICANT: Nabel, Gary J.  
APPLICANT: Gray, Gary S.  
APPLICANT: Remmert, Paul D.  
TITLE OF INVENTION: Methods For Selectively Stimulating Proliferation Of T-Cells  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSER: LAHIVE & COCKFIELD  
STREET: 60 State Street, Suite 510  
CITY: Boston



```

STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,816A
FILING DATE: May 4, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,253
FILING DATE: 10 MARCH 1995
APPLICATION NUMBER: US 08/253,964
FILING DATE: 3 JUNE 1994
APPLICATION NUMBER: US 08/073,223
FILING DATE: 4 JUNE 1993
APPLICATION NUMBER: US 08/200,947
FILING DATE: 23 FEB 1994
APPLICATION NUMBER: US 07/864,805
FILING DATE: 7 APR 1992
APPLICATION NUMBER: US 08/247,505
FILING DATE: 23 MAY 1994
APPLICATION NUMBER: US 07/864,866
FILING DATE: 7 APR 1992
APPLICATION NUMBER: US 08/218,155
FILING DATE: 25 MAR 1994
APPLICATION NUMBER: US 07/864,807
FILING DATE: 7 APR 1992
APPLICATION NUMBER: US 07/902,467
FILING DATE: 16 JUNE 1992
APPLICATION NUMBER: US 07/275,433
FILING DATE: 23 NOV 1988
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy B.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: RPT-002CP3
ELECTRONIC COMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 227 amino acids
TYPE: amino acid
TOPOLOGY: linear
COLLECTOR TYPE: protein
435-816A-6

Y Match 43.9%; Score 518.5; DB 4; Length 227;
Local Similarity 44.9%; Pred. No. 6.8e-52;
Res 101; Conservative 39; Mismatches 72; Indels 13; Gaps 5

1 KYILFVNFMVNLGAGVILGVALMLRHDPQTSLSLYLEGNKRAPIFTYGLIYLAVGA 60
  KYILFVNFMVNLGAGVILGVALMLRHDPQTSLSLYLEGNKRAPIFTYGLIYLAVGA 60
10 KYILFVNFMVNLGAGVILGVALMLRHDPQTSLSLYLEGNKRAPIFTYGLIYLAVGA 67
  KYILFVNFMVNLGAGVILGVALMLRHDPQTSLSLYLEGNKRAPIFTYGLIYLAVGA 67
61 VMNFVGLFCYGNIDIOSQCLSTFTFCVILIFACEVAGIGWFGVNRKQIALKDVKCPYDA 120
  VMNFVGLFCYGNIDIOSQCLSTFTFCVILIFACEVAGIGWFGVNRKQIALKDVKCPYDA 120
68 LMLVGLVGLFCYGNIDIOSQCLSTFTFCVILIFACEVAGIGWFGVNRKQIALKDVKCPYDA 127
  LMLVGLVGLFCYGNIDIOSQCLSTFTFCVILIFACEVAGIGWFGVNRKQIALKDVKCPYDA 127
121 LQAAVMDDDANNAKAVVKTFFETLNCSSNALITVLTITLR--NSLCPSSGNILPPLAQ 178
  LQAAVMDDDANNAKAVVKTFFETLNCSSNALITVLTITLR--NSLCPSSGNILPPLAQ 178
128 YNDLTKTKDPEQ--RETLAIHYALNCCG-----LAQGVROPTSDICPK-KVLELFTVK 178
  YNDLTKTKDPEQ--RETLAIHYALNCCG-----LAQGVROPTSDICPK-KVLELFTVK 178
179 DCHQKIDELFSGLYILGIALVAVVIMIFEMILSYLCCGIRNS 223
  DCHQKIDELFSGLYILGIALVAVVIMIFEMILSYLCCGIRNS 223
179 SCDDAIKEVFNKFIHIGAVGIVAVVIMIFEMILSYLCCGIRNS 223
  SCDDAIKEVFNKFIHIGAVGIVAVVIMIFEMILSYLCCGIRNS 223

```

```

: Sequence 1, Application US/08408222B
: Patent No. 5776727
:
: GENERAL INFORMATION:
: APPLICANT: Ikeyama, Shuichi
: APPLICANT: Koyama, Masaru
: APPLICANT: Miyake, Masayuki
: APPLICANT: Senoo, Masaharu
: TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dike, Bronstein, Roberts & Cushman
: STREET: 130 Water Street
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: PastSeq Version 1.5
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/408,222B
: FILING DATE: 22-MAR-1995
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/254,493
: FILING DATE: 06-JUN-1994
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP-079996-1991
: FILING DATE: 12-APR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP-085396-1991
: FILING DATE: 14-APR-1991
:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP-022321-1992
: FILING DATE: 07-FEB-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Resnick, David S.
: REGISTRATION NUMBER: 34,235
: REFERENCE/DOCKET NUMBER: 41777-DIV
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-523-3400
: TELEFAX: 617-523-6440
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 228 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: internal
: ORIGINAL SOURCE:
:
: US-08-408-222B-1
:
: Query Match 43.9%; Score 518.5; DB 1; Length 228;
: Best Local Similarity 44.9%; Pred. No. 6,8e-52;
: Matches 101; Conservative 39; Mismatches 72; Indels 13; Gaps 5;
:
: Oy 1 KTLFVFNFFVLAGVYILGVALMLRHPDPTTSLLYLELGNKKAAPTFFYGIYILIAVGA 60
: ||||| : : : : : : : : : : : : : : : : : : : : : : : :
: Db 11 KYLLFGNFFPLFAGLAVIALIGMLNFDSTKSIIFQETNN--NSSFFYGVYILIGAGA 68
:
: Oy 61 VMHVEVGLTGVGAIQISQCLLGTFFCTLVILFPCENVAGIWGFPVNDKQIAKDWQGFDOA 120
: :|| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
: Db 69 LMLVLGFLGCGAVQSSQCLGLFPGELVIPAIRIDAAIWAWSHDEVIKEVOEFPKDT 128
:
: Oy 121 LQGVAVDDDAANNAKAVKTFHRTLANCGSNALTTLTITLIR--NSLCPSCGNILTPILQ 178
: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
: Db 129 YNKLKTKRDBQ--RETLKAIVNALNCG-----LAGGVQRFIDSLCPK-KDVLFFFTVK 179

```

179 DCHOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGIRNS 223  
 180 SCPDAIKETVFNKPHIGAVGIGIAVIMIFEMILSMVLCGIRN 224

LT 7  
 8-808-148-3  
 quence 3, Application US/08808148  
 tent No. 6020478  
 GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer  
 APPLICANT: Goll, Surya  
 APPLICANT: Zhang, Hong Wolfe  
 TITLE OF INVENTION: NOVEL HUMAN TUMOR-ASSOCIATED ANTIGEN  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:

ADDRESSER: Incyte Pharmaceuticals, Inc.  
 STREET: 3174 Porter Drive  
 CITY: Palo Alto  
 STATE: CA

COUNTRY: USA  
 ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/808.148  
 FILING DATE: Herewith

CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
 REGISTRATION NUMBER: 36,749  
 REFERENCE/DOCKET NUMBER: PF-0218 US  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555  
 TELEFAX: 415-845-4166  
 TELEX:

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 237 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 180926

1-808-148-3

Query Match 25.3%; Score 298; DB 3; Length 237;  
 Local Similarity 30.7%; Pred. No. 2.3e-26;

Matches 75; Conservative 45; Mismatches 86; Indels 38; Gaps 5;

1 KYLLFVNFVMTAGVYILGVALMLRHPDPTTSLYLELNKRPAPNTFYVGIYILVAVGA 60  
 9 KYSMFNFPLFMICGILILALAIWVVSNDQAI---FCSBDVGSSTYAVADILLVAVGA 64  
 61 VMFVGLGCGYAIOSGCLLGTFTCLVILFACVVAAGTWGPNQDIADVKQ--FYD 118  
 65 IIMTLGLGCGGAIKESRCMLLPFGLILLILQVAVTGILGVFKSKSDRIWETLYEN 124  
 119 QALQAVMDDANNAKAVVKTFFHETLNCSSNALTTTLTILNSLCPSGNLI----- 172  
 125 TRLSANGSEKQFOEHLI-VFOBEFCCG-----LNGAADWGNNOHPPELCC 172  
 173 -----TPLLQODCHOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCG 218  
 173 ACIDKORPCOSYNGKQVYKETSIFIKDPLAKNLIIVIGISFGIAVEIIGLVFSMWLYC 232

QY 219 GIRN 222  
 DB 233 QIGN 236

RESULT 8

US-08-855-140-4  
 Sequence 4, Application US/08855140  
 Patent No. 5854022  
 GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.  
 APPLICANT: Bandman, Olga  
 APPLICANT: Goll, Surya K.  
 APPLICANT: Guegler, Karl J.  
 TITLE OF INVENTION: NEW TRANSMEMBRANE PROTEIN  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:

ADDRESSER: Incyte Pharmaceuticals, Inc.  
 STREET: 3174 Porter Drive  
 CITY: Palo Alto  
 STATE: CA

COUNTRY: USA  
 ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/855.140  
 FILING DATE: Herewith

CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
 REGISTRATION NUMBER: 36,749  
 REFERENCE/DOCKET NUMBER: PF-0296 US  
 TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555  
 TELEFAX: 415-845-4166  
 TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 219 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

LIBRARY: GenBank

CLONE: 1279546

US-08-855-140-4

Query Match 25.2%; Score 297.5; DB 2; Length 219;  
 Local Similarity 33.0%; Pred. No. 2.3e-26;

Matches 76; Conservative 38; Mismatches 85; Indels 31; Gaps 6;

1 KYLLFVNFVMTAGVYILGVALMLRHPDPTT-----SLYLELNKRPAPNTFYVGIYI 54  
 10 KYLLFVNFPLFMICGCGCIIIGFGLVPL-VQNTYGVLFRLMIFLITGN-----I 55  
 55 LIAVGVMMFVGLGCGYAIOSGCLLGTFTCLVILFACVVAAGTWGPNQDIADVKQ 114  
 56 LVIVGSIINWVAFGLGMSIKENKCLNSFVLLITILAEVYIALILFVYRQKNTLVA 115  
 115 QFYDQALQAVMDDANNAKAVVKTFFHETLNCSSNALTTTLTILNSLCPSGNLI 174  
 116 EGLNDSIQYHSDNSTMKAMDPIQT---QLCCGVNGSSDWTSG--PSSCSGADV--- 167  
 175 LLOODCHOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGCGIRNS 224  
 168 ---QGCYNFAKSMFHSNPLFYIGITICVCVIOVLGMSFALTLNCOIDKTS 214

1 9  
-430-225A-20  
Jence 20, Application US/08430225A  
ent No. 6204000  
GENERAL INFORMATION:  
APPLICANT: Dong, Jin-Tang; Barrett,  
APPLICANT: J. Carl; Lamb, Patricia W.; Isaacs, John T.  
TITLE OF INVENTION: DIAGNOSTIC METHODS AND  
TITLE OF INVENTION: GENE THERAPY USING REAGENTS DERIVED FROM THE  
TITLE OF INVENTION: HUMAN METASTASIS SUPPRESSOR GENE KAI1  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.  
STREET: 345 PARK AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: FLOPPY DISK  
COMPUTER: IBM PC COMPATIBLE  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WORDPERFECT 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/430,225A  
FILING DATE: 28-APR-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: RICHARD W. BORK  
REGISTRATION NUMBER: 36,459  
REFERENCE/DOCKET NUMBER: -2026-4172  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 758-4800  
TELEFAX: (212) 751-6849  
TELEX: 421792  
FORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 267 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
-430-225A-20

Y Match 23.1%; Score 272.5; DB 3; Length 267;  
Local Similarity 26.3%; Pred. No. 2.4e-23;  
hes 68; Conservative 50; Mismatches 84; Indels 57; Gaps 7;

1 KYLFVFNFFVFWLAGVILGVALLMRHPQT-TSLLYLKGKPKAPNTFYVGIYILI 59  
10 KYLFVFNFFVFWLAGVILGVALLMRHPQT-TSLLYLKGKPKAPNTFYVGIYILI 59  
60 AVMMFVFGVGLGCGVGAIOESQCLIGTEFTCLVILFACGVAAGVGFVFNKQDIADVKQF 119  
64 AVMMFVFGVGLGCGVGAIOESQCLIGTEFTCLVILFACGVAAGVGFVFNKQDIADVKQF 119  
120 ALQ--QAVVDDNANNAKAVVKTFFETLNCG----- 148  
124 LRDVNSSEDSIQDAMDVYA---QVKCCGWSFTNMTDNLAKMRPEVTPPCSEVAG 180  
149 --SNALTTTLTTLNLSLCPGSGNLT-----DLQODCHQKIDELFSGKLYLIGIA 198  
181 EEDNSLS-----VRKGCEAPENRTQSGNHPEDMDVYQEGCKEKVQAMIGENLGIILGV 234  
199 AIVAVIIMIFEMILSVLC 217  
235 GVGVALIELKAVLSTICLC 253

Patent No. 5854022  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K.  
APPLICANT: Guegler, Karl J.  
TITLE OF INVENTION: NEW TRANSMEMBRANE PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/855,140  
FILING DATE: Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0296 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
FORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 219 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 180141  
US-08-855-140-3

Query Match 22.5%; Score 265.5; DB 2; Length 219;  
Best Local Similarity 30.7%; Pred. No. 1.2e-22;  
Matches 70; Conservative 38; Mismatches 93; Indels 27; Gaps 5;

1 KYLFVFNFFVFWLAGVILGVALLMRHPQTSLY-----LELGNKPKAPNTFYVGIYILI 56  
10 KYLFVFNFFVFWLAGVILGVALLMRHPQTSLY-----LELGNKPKAPNTFYVGIYILI 56  
57 AVGAVMMFVFGVGLGCGVGAIOESQCLIGTEFTCLVILFACGVAAGVGFVFNKQDIADVKQF 116  
58 IGVSIIMVAVFLGCGVSIKKNKCLMSFILLIILLAEVTLAILFYEQGLNRYVAKG 117  
117 YDQALQAVVDDNANNAKAVVKTFFETLNCGSGNALTTTLTTLNLSLCPGSGNLTPL 176  
118 LTDSIHRYSN-----NSTYAADSDIOSFLQCCGINGTSDMTSG--PPACSPDRKV----- 167  
117 OODCHQKIDELFSGKLYLIGIAIYVAVIIMIFEMILSVLCGIRNS 224  
168 EGYAKRLMFSNFLTITGIIITTCVIEVAGSFAILLNCOIDKTS 214

RESULT 11  
US-08-807-044-3  
Sequence 3, Application US/08807044  
Patent No. 5863735  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.

[illegible]



123 LHIYGTQANVGLTNANSIIQT---DFRCGVSNYTDWFEVYNATRVVPDSCCLEFSSECGU 179  
172 LPELLOQDCHQKIDELFSGRUYLIGIAIVAVIMIPEMILSMVLCCGIRNSSVY 226  
180 HAPALVEGRATRGEGVASGELLAVGIFGLCTALVOLIGLINFAMTMTWQVVKADTY 234

ch completed: December 9, 2003, 10:33:33  
time : 31 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compen Ltd.

rotein - protein search, using sw model

on: December 9, 2003, 10:32:07 ; Search time 350 Seconds

(without alignments)  
120.032 Million cell updates/sec

ect score: US-10-035-914-2

ence: 1 KYLLFVNFVFWLAGVILG.....IFEMILSMVLCGGIRNSSVY 226

ing table: BLOSUM62

Gapop 10.0 , Gapext 0.5

ched: 684280 seqs, 185983659 residues

1 number of hits satisfying chosen parameters: 684280

mm DB seq length: 0

mm DB seq length: 200000000

-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

base :  
1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppaa/PTCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/PTCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query	Match	Length	ID	Description
1	1103	93.5	236	11	US-09-999-121-5 Sequence 5, Appli
2	1103	93.5	236	12	US-10-169-297-7 Sequence 7, Appli
3	1034	87.6	262	9	US-09-925-299-794 Sequence 794, App
4	1034	87.6	262	11	US-09-925-299-794 Sequence 794, App
5	1034	87.6	279	10	US-09-731-872-396 Sequence 396, App
6	1034	87.6	279	12	US-09-876-997-396 Sequence 396, App
7	522.5	44.3	226	12	US-10-205-194-162 Sequence 162, App
8	518.5	43.9	227	8	US-08-592-711-6 Sequence 6, Appli
9	518.5	43.9	227	11	US-09-350-202-6 Sequence 6, Appli
0	518.5	43.9	228	16	US-10-156-136-39 Sequence 39, Appl
1	518.5	43.9	275	9	US-09-925-301-1381 Sequence 1381, Ap
2	518.5	43.9	275	15	US-10-106-698-5930 Sequence 5930, Ap
3	462.5	39.2	221	12	US-10-205-219-185 Sequence 185, App
4	453.5	38.4	221	9	US-09-823-356-17 Sequence 17, Appl
5	453.5	38.4	221	16	US-10-156-136-25 Sequence 25, Appl

16	453.5	38.4	229	15	US-10-106-698-4519 Sequence 4519, Ap
17	400	33.9	209	15	US-10-106-698-6825 Sequence 6825, Ap
18	339	28.7	157	11	US-09-823-187-76 Sequence 76, Appl
19	316	26.8	254	10	US-09-934-268-4 Sequence 4, Appli
20	316	26.8	254	15	US-10-162-435-38 Sequence 38, Appl
21	298	25.3	237	9	US-09-922-217-1062 Sequence 1062, Ap
22	298	25.3	237	10	US-09-833-263-1062 Sequence 1062, Ap
23	298	25.3	237	14	US-10-025-380-1062 Sequence 1062, Ap
24	298	25.3	241	9	US-09-925-301-1035 Sequence 1035, Ap
25	298	25.3	241	15	US-10-106-698-5105 Sequence 5105, Ap
26	298	25.3	243	9	US-09-922-217-1122 Sequence 1122, Ap
27	298	25.3	243	14	US-10-025-380-1122 Sequence 1122, Ap
28	298	25.3	446	9	US-09-922-217-1121 Sequence 1121, Ap
29	298	25.3	446	14	US-10-025-380-1121 Sequence 1121, Ap
30	297.5	25.2	219	10	US-09-976-782-118 Sequence 118, App
31	289.5	24.5	249	10	US-09-908-193-16 Sequence 16, Appl
32	288	24.4	248	10	US-09-908-193-41 Sequence 41, Appl
33	288	24.4	248	15	US-10-103-196-15 Sequence 15, Appl
34	286.5	24.3	219	12	US-09-976-782-117 Sequence 117, App
35	286.5	24.3	219	16	US-10-156-136-34 Sequence 34, Appl
36	283.5	24.0	247	10	US-09-908-193-42 Sequence 42, Appl
37	283.5	24.0	247	11	US-09-823-187-72 Sequence 72, Appl
38	282.5	23.9	239	9	US-09-823-356-14 Sequence 14, Appl
39	272.5	23.1	239	12	US-09-976-782-119 Sequence 119, App
40	272.5	23.1	267	9	US-09-785-380-20 Sequence 20, Appl
41	267.5	22.7	267	11	US-09-823-187-73 Sequence 73, Appl
42	265.5	22.5	219	11	US-09-823-187-75 Sequence 75, Appl
43	265.5	22.5	219	12	US-09-970-424-3 Sequence 3, Appli
44	265.5	22.5	219	12	US-09-976-782-116 Sequence 116, App
45	265.5	22.5	231	9	US-09-925-302-474 Sequence 474, App

ALIGNMENTS

RESULT 1  
US-09-999-121-5  
; Sequence 5, Application US/09999121  
; Publication No. US20030039982A1  
; GENERAL INFORMATION:  
; APPLICANT: Ryan, James W.  
; TITLE OF INVENTION: ISOLATED GENOMIC NUCLEOTIDE FRAGMENTS FROM THE p15 REGION OF  
; FILE REFERENCE: CHROMOSOME 11  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: US/09/999,121  
; PRIOR FILING DATE: 2000-10-31  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5  
; LENGTH: 236  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-999-121-5

Query Match 93.5%; Score 1103; DB 11; Length 236;  
Best Local Similarity 91.6%; Pred. No. 1.6e-107;

Matches 207; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

QY 1 KYLLFVNFVFWLAGVILGVALMLRHPDQTSLVILBLGNKPAFTFYGIYILAVGA 60  
DB 11 KYLLFVNFVFWLAGVILGVALMLRHPDQTSLVILBLGDKRAPFTFYGIYILAVGA 70  
QY 61 VMWFVFLGCTGAIOSQCLGFFFCVLVLPACRYAAGIWKGVNMDQIAKYKOTYDQA 120  
DB 71 VMWFVFLGCTGAIOSQCLGFFFCVLVLPACRYAAGIWKGVNMDQIAKYKOTYDQA 130  
QY 121 LQAAVDDDDANNAKAAVVKTFHETLNCSSNALTTTLTILRNLSLCPSGNIILTPLLQOOC 180  
DB 131 LQAAVDDDDANNAKAAVVKTFHETLNCSSNALTTTLTILRNLSLCPSGNIILNKREDC 190  
QY 181 HOKIDBLFSGKYLITGLAIIVAVVIMIFEMILSMVLCGGIRNSSVY 226

|||||  
191 HOKIDDLFSGKLYLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 236

T 2  
-169-297-7  
Application US/10169297  
Sequence 7, Application US/10169297  
Patent No. US20030171276A1  
Patent Information:

PLICANT: Morata, Naoki

PLICANT: Enjoji, Takashi

TITLE OF INVENTION: Preventives and Remedies for Chronic

HEPATITIS

FILE OF INVENTION: Hepatitis

RE REFERENCE: 3435.1000-000

RENT APPLICATION NUMBER: US/10/169,297

RENT FILING DATE: 2002-10-31

FOR APPLICATION NUMBER: PCT/JP00/09393

FOR FILING DATE: 2000-12-28

FOR APPLICATION NUMBER: JP 11/374087

FOR FILING DATE: 1999-12-28

MEMBER OF SEQ ID NOS: 50

FTWABE: FastSeq for Windows Version 4.0

ID NO 7

ENGTH: 236

YPR: PRT

RGANISM: Homo sapiens

-169-297-7

Match

Local Similarity 93.5%; Score 1103; DB 12; Length 236;

Local Similarity 91.6%; Pred. No. 1.6e-107;

Matches 207; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

1 KTLFVFNVLAVAGVILGVALMLRHPDPTSLYLKLNKPAVNTFYGIYLLAVGA 60

11 KTLFVFNVLAVAGVILGVALMLRHPDPTSLYLKLNKPAVNTFYGIYLLAVGA 70

61 VMMFVGLGCTGALIGSCLLGFPTCLVILPACVNAAGIMGFVNDQIAKQKPYDQA 120

71 VMMFVGLGCTGALIGSCLLGFPTCLVILPACVNAAGIMGFVNDQIAKQKPYDQA 130

121 LQAAVDDDDANNAKAVVKTFFHTLNCSSNALTTTLIRNSLCPSGNIPLPQLQDC 180

131 LQAAVDDDDANNAKAVVKTFFHTLNCSSNALTTTLIRNSLCPSGNIPLPQLQDC 190

181 HOKIDDLFSGKLYLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 226

191 HOKIDDLFSGKLYLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 236

T 3

-925-299-794

ence 794, Application US/09925299

Pat No. US20020055627A1

Patent Information:

PLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

RE REFERENCE: PA102

RENT APPLICATION NUMBER: US/09/925,299

RENT FILING DATE: 2001-08-10

FOR APPLICATION NUMBER: PCT/US00/05883

FOR FILING DATE: 2000-03-08

FOR APPLICATION NUMBER: 60/124,270

FOR FILING DATE: 1999-03-12

MEMBER OF SEQ ID NOS: 1556

FTWABE: Patent In Ver. 2.0

ID NO 794

ENGTH: 262

YPR: PRT

RGANISM: Homo sapiens

ZATURE: Homo sapiens

ME/KEY: SITE

CATION: (38)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-794

Query Match 87.6%; Score 1034; DB 9; Length 262;  
Best Local Similarity 91.1%; Pred. No. 3.3e-100;  
Matches 195; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

13 LAGVILGVALMLRHPDPTSLYLKLNKPAVNTFYGIYLLAVGAVMMFVGLGCTG 72

49 LAGVILGVALMLRHPDPTSLYLKLNKPAVNTFYGIYLLAVGAVMMFVGLGCTG 108

73 AIOESQCLLGFPTCLVILPACVNAAGIMGFVNDQIAKQKPYDQAALQAAVDDDDANN 132

109 AIOESQCLLGFPTCLVILPACVNAAGIMGFVNDQIAKQKPYDQAALQAAVDDDDANN 168

133 AKAVKTFFHTLNCSSNALTTTLIRNSLCPSGNIPLPQLQDCQKIDDLFSGKL 192

169 AKAVKTFFHTLNCSSNALTTTLIRNSLCPSGNIPLPQLQDCQKIDDLFSGKL 228

193 YLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 226

229 YLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 262

RESULT 4

US-09-925-299-794

Sequence 794, Application US/09925299

Publication No. US20030040617A9

GENERAL INFORMATION:

PLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA102

CURRENT APPLICATION NUMBER: US/09/925,299

CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05883

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 1556

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 794

LENGTH: 262

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (38)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-299-794

Query Match 87.6%; Score 1034; DB 11; Length 262;

Best Local Similarity 91.1%; Pred. No. 3.3e-100;

Matches 195; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

13 LAGVILGVALMLRHPDPTSLYLKLNKPAVNTFYGIYLLAVGAVMMFVGLGCTG 72

49 LAGVILGVALMLRHPDPTSLYLKLNKPAVNTFYGIYLLAVGAVMMFVGLGCTG 108

73 AIOESQCLLGFPTCLVILPACVNAAGIMGFVNDQIAKQKPYDQAALQAAVDDDDANN 132

109 AIOESQCLLGFPTCLVILPACVNAAGIMGFVNDQIAKQKPYDQAALQAAVDDDDANN 168

133 AKAVKTFFHTLNCSSNALTTTLIRNSLCPSGNIPLPQLQDCQKIDDLFSGKL 192

169 AKAVKTFFHTLNCSSNALTTTLIRNSLCPSGNIPLPQLQDCQKIDDLFSGKL 228

193 YLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 226

229 YLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 262

RESULT 5

US-09-731-872-396



Sequence 396, Application US/09731872  
 Patent No. US20020102604A1  
 GENERAL INFORMATION:  
 APPLICANT: Dumas Milne Edwards, Jean Baptiste  
 APPLICANT: Bougueleret, Lydie  
 APPLICANT: Jobert, Severin  
 TITLE OF INVENTION: FULL-LENGTH HUMAN CDNA5 ENCODING POTENTIALLY SECRETED PROTEINS  
 FILE REFERENCE: 78.053.REG  
 CURRENT APPLICATION NUMBER: US/09/731.872  
 PRIOR FILING DATE: 2000-12-07  
 PRIOR APPLICATION NUMBER: US 60/169,629  
 PRIOR FILING DATE: 1999-12-08  
 PRIOR APPLICATION NUMBER: US 60/187,470  
 PRIOR FILING DATE: 2000-03-06  
 NUMBER OF SEQ ID NOS: 482  
 SOFTWARE: Patent.pm  
 Q ID NO 396  
 LENGTH: 279  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 9-731-872-396

Query Match 87.6%; Score 1034; DB 10; Length 279;  
 Local Similarity 91.1%; Pred. No. 3.5e-100;  
 Mismatches 14; Indels 0; Gaps 0;

13 LAGVILGVALLMLRHPDPTSLIYLELGNKRPAPNTFYVGIYLLIIVGAVMMVGVGLGCG 72  
 66 LAGVILGVALLMLRHPDPTSLIYLELGNKRPAPNTFYVGIYLLIIVGAVMMVGVGLGCG 125  
 73 AIOESQCLGTFCTCLVILFACVAAAGIWMGFVFNKQIADVQKQFDQALQQAAMDDANN 132  
 126 AIOESQCLGTFCTCLVILFACVAAAGIWMGFVFNKQIADVQKQFDQALQQAAMDDANN 185  
 133 AKAVVKTETHTLNCSSNALITLTTILNLSLCPGSGNITLPLLODCHOKIDELFSGKL 192  
 186 AKAVVKTETHTLNCSSNALITLTTILNLSLCPGSGNITLPLLODCHOKIDELFSGKL 245  
 193 YLIGIAIVAVVIMIFEMILSMVLCGGIRNSSVY 226  
 246 YLIGIAIVAVVIMIFEMILSMVLCGGIRNSSVY 279

Query Match 87.6%; Score 1034; DB 12; Length 279;  
 Local Similarity 91.1%; Pred. No. 3.5e-100;  
 Mismatches 14; Indels 0; Gaps 0;

13 LAGVILGVALLMLRHPDPTSLIYLELGNKRPAPNTFYVGIYLLIIVGAVMMVGVGLGCG 72  
 66 LAGVILGVALLMLRHPDPTSLIYLELGNKRPAPNTFYVGIYLLIIVGAVMMVGVGLGCG 125  
 73 AIOESQCLGTFCTCLVILFACVAAAGIWMGFVFNKQIADVQKQFDQALQQAAMDDANN 132  
 126 AIOESQCLGTFCTCLVILFACVAAAGIWMGFVFNKQIADVQKQFDQALQQAAMDDANN 185  
 133 AKAVVKTETHTLNCSSNALITLTTILNLSLCPGSGNITLPLLODCHOKIDELFSGKL 192  
 186 AKAVVKTETHTLNCSSNALITLTTILNLSLCPGSGNITLPLLODCHOKIDELFSGKL 245  
 193 YLIGIAIVAVVIMIFEMILSMVLCGGIRNSSVY 226  
 246 YLIGIAIVAVVIMIFEMILSMVLCGGIRNSSVY 279

Query Match 44.3%; Score 522.5; DB 12; Length 226;  
 Local Similarity 45.1%; Pred. No. 1.2e-46;  
 Mismatches 101; Indels 13; Gaps 5;

1 KYLFFVNFVFWLGLAGVILGVALLMLRHPDPTSLIYLELGNKRPAPNTFYVGIYLLIIVG 60  
 11 KYLFFVNFVFWLGLAGVILGVALLMLRHPDPTSLIYLELGNKRPAPNTFYVGIYLLIIVG 66  
 61 VMMFVGLGCGVGAIOESQCLGTFCTCLVILFACVAAAGIWMGFVFNKQIADVQKQFDQAL 120  
 67 LMMVGLGCGVGAIOESQCLGTFCTCLVILFACVAAAGIWMGFVFNKQIADVQKQFDQAL 126  
 121 LQQAAMDDANNNAKAVVKTETHTLNCSSNALITLTTILNLSLCPGSGNITLPLLOD 179  
 127 YOKLRNDEPO--RETUKAITHMALMCCGAGVGFQFISDI-----CRK-KQVLESFQVS 178  
 180 CHOKIDELFSGKLYLIGIAIVAVVIMIFEMILSMVLCGGIRNSSVY 223  
 179 CPDAIDVFSKTHITIGAVGIAVAVVIMIFEMILSMVLCGGIRNSSVY 222

Query Match 44.3%; Score 522.5; DB 12; Length 226;  
 Local Similarity 45.1%; Pred. No. 1.2e-46;  
 Mismatches 101; Indels 13; Gaps 5;

1 KYLFFVNFVFWLGLAGVILGVALLMLRHPDPTSLIYLELGNKRPAPNTFYVGIYLLIIVG 60  
 11 KYLFFVNFVFWLGLAGVILGVALLMLRHPDPTSLIYLELGNKRPAPNTFYVGIYLLIIVG 66  
 61 VMMFVGLGCGVGAIOESQCLGTFCTCLVILFACVAAAGIWMGFVFNKQIADVQKQFDQAL 120  
 67 LMMVGLGCGVGAIOESQCLGTFCTCLVILFACVAAAGIWMGFVFNKQIADVQKQFDQAL 126  
 121 LQQAAMDDANNNAKAVVKTETHTLNCSSNALITLTTILNLSLCPGSGNITLPLLOD 179  
 127 YOKLRNDEPO--RETUKAITHMALMCCGAGVGFQFISDI-----CRK-KQVLESFQVS 178  
 180 CHOKIDELFSGKLYLIGIAIVAVVIMIFEMILSMVLCGGIRNSSVY 223  
 179 CPDAIDVFSKTHITIGAVGIAVAVVIMIFEMILSMVLCGGIRNSSVY 222

51 apps

54





OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/156,136  
FILING DATE: 29-May-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/010,146  
FILING DATE: <Unknown>  
APPLICATION NUMBER: WO US98/00959  
FILING DATE: 21-JAN-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: A. Anders Brookes  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PF354PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 301-309-8504  
FAX: 301-309-8439  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 221 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 25:  
0-156-136-25

Seq Match 38.4%, Score 453.5, DB 16, Length 221,  
Local Similarity 41.3%, Pred. No. 2e-39,  
tches 92; Conservative 33; Mismatches 81; Indels 17; Gaps 4;

```
1 KILFVFNFMAGVILGVALMLRHPQTSILYLBIQNKPAFPTVYGIYILAVGA 60
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
12 KILLIGFNLTFILASNAVAFGLMFRFGAIKEL-----SSDKSPRYFVGLYLVAGGA 67
61 VMFVGFAGCYAIGESQCLGTFTCLVILFACVAAAGVGFVNNDQIAKDYKQFYDOA 120
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
68 LMAVAVGFPGCCGAMESQCVLSFTCLVIFAEVTVGVFAFVGKVAIRHVQMTYEA 127
121 LQQAAMDDANNAKAVKTFHETLNCSSNALTTTLIRNSLCPSGGNLTPLLODC 180
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
128 YNDYKDRGKNGTLI--TFHSTFQCCGKSSSEGVQPT-----CPK-----EILGHKNC 174
181 HOKIDELPSGKLYLIGIAIYAVAVMIFEMILSMTCGIRNS 223
: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
175 IDRIETIISVKQLIGIVGIGIAGLTFGMIFSMVLCGAIIRNS 217
```

Job completed: December 9, 2003, 11:10:20  
Time : 352 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comphen Ltd.

rotein - protein search, using sw model

on: December 9, 2003, 09:56:34 / Search time 36 Seconds  
(without alignments)  
603.726 Million cell updates/sec

ect score: 1180  
US-10-035-914-2

ence: 1 KYLLFVFNFWFLAGVILG.....IFEMILSMVLCGGRNSSVY 226

ing table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

ched: 283308 segs, 96168682 residues

l number of hits satisfying chosen parameters: 283308

num DB seq length: 0  
num DB seq length: 200000000

-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

ase :  
1: PIR.76:\*  
2: PIR1.\*  
3: PIR3.\*  
4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Score	Query Match	Length	ID	Description
1 1177	99.7	236	A46472	cell surface prote
2 1103	93.5	236	A35649	cell surface prote
3 528.5	44.8	226	UX0221	CD9 antigen - bovi
4 522.5	44.3	226	S39262	CD9 antigen - rat
5 518.5	43.9	228	A40402	CD9 antigen [valid
6 512.5	43.4	228	A42929	CD9 antigen - gree
7 510.5	43.3	226	I49589	antigen - mouse
8 420	35.6	222	A59253	tetraspan TSPAN-2
9 298	25.3	237	A36056	tumor-associated a
0 286.5	24.3	219	A39574	leukocyte antigen
1 272.5	23.1	267	A46493	metastasis suppres
2 265.5	22.5	219	A37243	hemopoietic cell s
3 262	22.2	266	I49561	C33/R2/IR4 - mouse
4 255.5	21.7	282	T21596	hypothetical prote
5 254.5	21.6	228	A59265	tetraspan TSPAN-4
6 227.5	19.3	233	T15620	hypothetical prote
7 221.5	18.5	241	A59262	tetraspan TSPAN-1
8 218.5	18.5	223	T26763	hypothetical prote
9 209	17.7	244	I39368	T-cell acute lymph
0 207	17.5	238	S43511	CD63/MR91 antigen
1 205	17.4	245	A59258	tetraspan TSPAN-3
2 199.5	16.9	254	T13615	hypothetical prote
3 196	16.6	253	A59264	tetraspan TSPAN-6
4 195	16.5	245	A46508	CD63/MR91 antigen
5 194.5	16.5	281	A47629	cell surface glyco
6 191	16.2	264	A59261	tetraspan TSPAN-5
7 187	15.8	238	UC2297	CD63 antigen - rab
8 180.5	15.3	281	B47629	cell surface glyco

30	178.5	15.1	308	2	T24912	hypothetical prote
31	178	15.1	242	2	T15361	hypothetical prote
32	174.5	14.8	206	2	T25161	hypothetical prote
33	170	14.4	218	1	A43522	23k integral membr
34	170	14.4	238	1	I38016	melanoma - associa
35	162	13.7	321	2	T45053	hypothetical prote
36	158	13.4	218	1	A40181	hypothetical prote
37	149.5	12.7	194	2	T25548	hypothetical prote
38	141.5	12.0	394	2	T18752	hypothetical prote
39	113	9.6	203	2	T22537	hypothetical prote
40	113	9.6	263	2	T02906	senescence-associa
41	109	9.2	427	2	T32652	hypothetical prote
42	106	9.0	244	2	S44610	C02F5.8 protein -
43	105	8.9	260	2	T46081	uroplakin Ib - bov
44	100.5	8.5	359	2	T18667	hypothetical prote
45	100	8.5	258	2	I46080	uroplakin Ia - bov

ALIGNMENTS

RESULT 1  
A46472  
cell surface protein TAPA-1 - mouse  
N/Alternate names: target of antiProliferative antibody (TAPA) 1  
C/Species: Mus musculus (house mouse)  
C/Date: 18-Jun-1993 #sequence\_revision 09-Aug-1996 #text\_change 22-Jun-1999  
R/Andrila, M.L.; Hsieh, C.L.; Oren, R.; Francke, U.; Levy, S.  
J. Immunol. 147, 1030-1036, 1991  
A/Title: Genomic organization and chromosomal localization of the TAPA-1 gene.  
A/Reference number: A46472; MUID:91318144; PMID:1650385  
A/Accession: A46472  
A/Molecule type: DNA  
A/Residues: 1-236 <AND>  
A/Cross-references: GB:845012; NID:91679982; PIDN:RA819417.1; PID:9233253  
A/Experimental source: B-cell lymphoma line 38C13  
A/Note: Sequence extracted from NCBI backbone (NCBIN:44957, NCBIN:44966, NCBIN:45001, NCBI  
C/Superfamily: CD9 antigen  
C/Keywords: transmembrane protein  
F/2-11/Domain: intracellular #status predicted <CY1>  
F/12-35/Domain: transmembrane #status predicted <TM1>  
F/36-57/Domain: extracellular #status predicted <EX1>  
F/58-84/Domain: transmembrane #status predicted <TM2>  
F/85-113/Domain: intracellular #status predicted <CY2>  
F/89-113/Domain: transmembrane #status predicted <TM3>  
F/114-203/Domain: extracellular #status predicted <EX2>  
F/204-230/Domain: transmembrane #status predicted <TM4>  
F/231-236/Domain: intracellular #status predicted <CY3>

Query Match 99.7%; Score 1177; DB 1; Length 236;  
Best Local Similarity 99.6%; Pred. 1e-97; Indels 0; Gaps 0;  
Matches 225; Conservative 1; Mismatches 0;

QY	1 KYLLFVFNFWFLAGVILGVALMLRHPDQTSLSLYLBLEGNRPAPTFYVGIIYLLAVGA 60
DB	11 KYLLFVFNFWFLAGVILGVALMLRHPDQTSLSLYLBLEGNRPAPTFYVGIIYLLAVGA 70
QY	61 VMWFGFLGCGAIOBSOCLLGTFFPCVLVILFACRFAVAGINQFPAKQOIAKYOKFYDOA 120
DB	71 VMWFGFLGCGAIOBSOCLLGTFFPCVLVILFACRFAVAGINQFPAKQOIAKYOKFYDOA 130
QY	121 LQOAVWDDDANNAKAVKTFHETLNCSSNALTTTLTLRSLCPGSGNITLPLLOQOC 180
DB	131 LQOAVWDDDANNAKAVKTFHETLNCSSNALTTTLTLRSLCPGSGNITLPLLOQOC 190
QY	181 HOKIDELPSGKLYLIGIAIVAVIMIFEMILSMVLCGGRNSSVY 226
DB	191 HOKIDELPSGKLYLIGIAIVAVIMIFEMILSMVLCGGRNSSVY 236
RESULT 2	A35649



180 CHQKIDELFSGKLYLIGIAIVAVMIFEMILSMVLCGGRNS 223  
 179 CPDAIDVFPKSHHIIIGAVGIGIAVVMIFEMIFSMILCCAIRRS 222

LT 5  
 02 antigen [validated] - human  
 alternate names: motility-related protein-1  
 ecies: Homo sapiens (man)  
 ce: 06-Dec-1991 #sequence revision 07-Jul-1995 #text change 08-Dec-2000  
 bnfstein, E.; Benoit, P.; Billard, M.; Plaisance, S.; Prenant, M.; Uzan, G.; Bouchet  
 mcs 16, 132-138, 1993  
 le: Organization of the human CD9 gene.  
 ference number: A46123; MUID:93252369; PMID:8486348  
 session: A46123

le: CDNA cloning and expression of platelet p24/CD9. Evidence for a new family of m  
 session: A40402; MUID:91244846; PMID:2037603  
 ference number: A40402  
 itus: not compared with conceptual translation  
 le: not compared with conceptual translation  
 idues: 1-228 <LAN>  
 88-references: GB:J34068; GB:M61880; NID:G508495; PIDN:AAA59982.1; PID:G508496  
 e: parts of this sequence, including the amino end of the mature protein, were conf  
 ke, M.; Koyama, M.; Seno, M.; Ikegama, S.  
 p. Med. 174, 1347-1354, 1991  
 le: Identification of the motility-related protein (MRP-1), recognized by monoclonal  
 session: JH0555; MUID:92078843; PMID:1720807

le: not compared with conceptual translation  
 idues: 1-228 <MTY>  
 88-references: GB:X60111; NID:G34768; PIDN:CAA2708.1; PID:G34769  
 erimental source: breast carcinoma  
 e: this protein has the epitope defined by cell motility-inhibiting monoclonal anti  
 chelx, C.; Benoit, P.; Prachet, P.; Billard, M.; Worthington, R.E.; Gagnon, J.; Uza  
 ol. Chem. 266, 117-122, 1991  
 le: Molecular cloning of the CD9 antigen. A new family of cell surface proteins.  
 ference number: A39029; MUID:91093112; PMID:1840589  
 session: A39029

le: not compared with conceptual translation  
 idues: 1-8, 'S', '10-66, 'A', '68-193, '195-228 <BOU>  
 88-references: GB:M38690  
 e: parts of this sequence, including the amino end of the mature protein, were conf  
 ashihara, M.; Takahata, K.; Yatomi, Y.; Nakahata, K.; Kurokawa, K.  
 Lett. 264, 270-274, 1990  
 le: Purification and partial characterization of CD9 antigen of human platelets.  
 erence number: S10564; MUID:90292223; PMID:2358073  
 session: S10564

le: not compared with conceptual translation  
 idues: 2-8, 'X', '10-21 <HIG>  
 e: GDB:CD9; MICS  
 88-references: GDB:120582; OMTM:143030  
 position: 12P13-12P13  
 exfamily: CD9 antigen  
 words: glycoprotein; transmembrane protein  
 26/Product: CD9 antigen #status experimental <MAT>  
 1/Domain: intracellular #status predicted <CY1>  
 35/Domain: transmembrane #status predicted <EX1>  
 55/Domain: extracellular #status predicted <EX1>  
 82/Domain: transmembrane #status predicted <TM2>  
 86/Domain: intracellular #status predicted <CY2>  
 111/Domain: transmembrane #status predicted <TM3>  
 -194/Domain: extracellular #status predicted <EX2>

F:195-221/Domain: transmembrane #status predicted <TM4>  
 F:222-228/Domain: intracellular #status predicted <CY3>  
 F:53/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 43.9%; Score 518.5; DB 1; Length 228;  
 Best Local Similarity 44.4%; Pred. No. 6.1e-39;  
 Matches 101; Conservative 39; Mismatches 72; Indels 13; Gaps 5;

QY 1 KYLLFVNFVFMAGVILGVALLMRHDPOTTSLLVLELGNKPAFPTFYGIYLLIANGA 60  
 Db 11 KYLLFVNFVFMAGVILGVALLMRHDPOTTSLLVLELGNKPAFPTFYGIYLLIANGA 68  
 QY 61 VMNFGFLGCGAIOESQCLGTFPTCLVILFACVAVGICFPNNQIOADVQKPYDQA 120  
 Db 69 LMLVGFILGCGAVQESQCLGTFPTCLVILFALBIAAIWGHKQVIREVOEFYKDT 128  
 QY 121 LQAAVMDDDANNAKAVKTFHETLNCSSNALTTTLTLR--NSLCPSGNITPLDQ 178  
 Db 129 YNKLKTRDEPQ--RETLKAIHYALNCCG-----LAGVRFQFSDICPK-KDVLFTPTVK 179  
 QY 179 DCHQKIDELFSGKLYLIGIAIVAVMIFEMILSMVLCGGRNS 223  
 Db 180 SCPDAIKKVFNNKHHIIIGAVGIGIAVVMIFEMIFSMILCCAIRRN 224

RESULT 6

A42929  
 CD9 antigen - green monkey  
 N:Alternate names: 27K diphtheria toxin receptor-associated protein DRAP27  
 C:Species: Cercopithecus aethiops (green monkey, grivet)  
 C:Date: 01-Oct-1992 #sequence revision 09-Aug-1996 #text change 16-Jun-2000  
 R:Accession: A42929  
 R:Mitsumura, T.; Yamoto, R.; Unata, T.; Yomo, T.; Urabe, I.; Tsunooka, M.; Mekada, E.  
 J. Cell Biol. 118, 1389-1399, 1992  
 A>Title: The 27-kD diphtheria toxin receptor-associated protein (DRAP27) from vero cells  
 eptors on toxin-sensitive cells.  
 A:Reference number: A42929; MUID:92394967; PMID:1522113  
 A:Accession: A42929  
 A>Status: nucleic acid sequence not shown  
 A:Molecule type: mRNA  
 A:Residues: 1-228 <MTY>  
 A:Cross-references: GB:D10726; NID:G218565; PIDN:BAA01569.1; PID:G218566  
 C:Superfamily: CD9 antigen  
 C:Keywords: glycoprotein; transmembrane protein  
 F:2-228/Product: CD9 antigen #status predicted <MAT>  
 F:2-11/Domain: intracellular #status predicted <CY1>  
 F:12-35/Domain: transmembrane #status predicted <EX1>  
 F:36-55/Domain: extracellular #status predicted <EX1>  
 F:56-82/Domain: transmembrane #status predicted <TM2>  
 F:83-86/Domain: intracellular #status predicted <CY2>  
 F:87-111/Domain: transmembrane #status predicted <TM3>  
 F:112-194/Domain: extracellular #status predicted <EX2>  
 F:195-221/Domain: transmembrane #status predicted <TM4>  
 F:222-228/Domain: intracellular #status predicted <CY3>  
 F:52/53/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 43.4%; Score 512.5; DB 1; Length 228;  
 Best Local Similarity 44.4%; Pred. No. 2.1e-38;  
 Matches 100; Conservative 40; Mismatches 72; Indels 13; Gaps 5;

QY 1 KYLLFVNFVFMAGVILGVALLMRHDPOTTSLLVLELGNKPAFPTFYGIYLLIANGA 60  
 Db 11 KYLLFVNFVFMAGVILGVALLMRHDPOTTSLLVLELGNKPAFPTFYGIYLLIANGA 68  
 QY 61 VMNFGFLGCGAIOESQCLGTFPTCLVILFACVAVGICFPNNQIOADVQKPYDQA 120  
 Db 69 LMLVGFILGCGAVQESQCLGTFPTCLVILFALBIAAIWGHKQVIREVOEFYKDT 128  
 QY 121 LQAAVMDDDANNAKAVKTFHETLNCSSNALTTTLTLR--NSLCPSGNITPLDQ 178  
 Db 129 YNKLKTRDEPQ--RETLKAIHYALNCCG-----LAGVRFQFSDICPK-KDVLFTPTVK 179  
 QY 179 DCHQKIDELFSGKLYLIGIAIVAVMIFEMILSMVLCGGRNS 223







family: CD9 antigen  
 order: glycoprotein; transmembrane protein  
 16/Domain: intracellular #status predicted <CY1>  
 15/Domain: transmembrane #status predicted <TM1>  
 14/Domain: extracellular #status predicted <EX1>  
 13/Domain: transmembrane #status predicted <TM2>  
 12/Domain: intracellular #status predicted <CY2>  
 11/Domain: transmembrane #status predicted <TM3>  
 10/Domain: extracellular #status predicted <EX2>  
 9/Domain: transmembrane #status predicted <TM4>  
 8/Domain: intracellular #status predicted <CY3>  
 7/Binding site: carbohydrate (Asn) (covalent) #status predicted

Match 22.5%; Score 265.5; DB 1; Length 219;  
 Local Similarity 30.7%; Pred. No. 2.3e-16;  
 Conservative 38; Mismatches 93; Indels 27; Gaps 5;

1 KYLLFVNFVFMALAGVILGVALMLRHPDPTSLYLTLTILNSLCPSGNILTPIL 56  
 10 KYLLFVNFVFMALAGVILGVALMLRHPDPTSLYLTLTILNSLCPSGNILTPIL 57  
 57 AVGVAMFVFGICGALIOESQCLGTFPTCLVILPACVAAAGVFNKQIADVQK 116  
 58 IYGSIMVAVFLGCMGSIKRNKCLMSFILLIILAEVTLAILFVYEQKLREYVAKG 117  
 117 YDQAQAAVMDANNAAKAVVTFHETLNCSSNALTTITLILNSLCPSGNILTPIL 176  
 118 LDTSHRHS---NSTKAMDSIOGFLCCGCGINGTSDMTSG--PPASCPDRKV----- 167

177 QODCHQKIDELFSKLYLIGIAIIVAVVIMIFEMILSMVLCCGIRNS 224  
 168 EGVAKARLNFHNSNLYIGITTCVCEVLEAGMSFALTLNCOIDKTS 214

13  
 114 - mouse  
 113  
 112 Mus musculus (house mouse)  
 111 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 20-Aug-1999  
 110 149561  
 109 ra, M.; Imai, T.; Ishikawa, I.; Uwabe, K.I.; Yoshie, O.  
 108 Immunol. 157, 144-157, 1994  
 107 e: Mouse homologue of C33 antigen (CD82), a member of the transmembrane 4 superfam  
 106 149561; MUID:9413678; PMID:8039242  
 105 149561  
 104 us: preliminary; translated from GB/EMBL/DBJ  
 103 cule type: mRNA  
 102 1-266 <RBS>  
 101 s-references: GB:D14883; NID:984036; PIDN:BAA03602.1; PID:9498368  
 100 rfamily: CD9 antigen

Match 22.2%; Score 262; DB 2; Length 266;  
 Local Similarity 27.2%; Pred. No. 5.7e-16;  
 Conservative 47; Mismatches 81; Indels 62; Gaps 10;

1 KYLLFVNFVFMALAGVILGVALMLRHPDPTSLYLTLTILNSLCPSGNILTPIL 59  
 10 KYLLFVNFVFMALAGVILGVALMLRHPDPTSLYLTLTILNSLCPSGNILTPIL 63  
 60 AVGVAMFVFGICGALIOESQCLGTFPTCLVILPACVAAAGVFNKQIADVQK 119  
 64 ATIYVGFGLGICGAVNRKCLGLFVFLILLIAGVTVGVLFYNADKLKRS----- 116  
 120 ALQQAQVMDANNAAKAVVTFHETLNCSSNALTTITLILNSLCPSGNILTPIL 158  
 117 -MANTVMDIIRYTNATSSREBANDYQAVKCGWVSHYMTNBEIEMGFTKTYPCS 175  
 159 -----ILNSLCPSGNILTPIL-----PLDQCHQK-----IDELFSKLYLIG 196  
 176 CEKIKEDNQILVKKGFCEADNSTYSNNPBMVPTVTEGCKEKAQAMLOENFG---ILIG 232  
 197 IAAIVAVVIMIFEMILSMVL 217

Db 233 VCA-GVAVTELLGLFLSLICLC 252

RESULT 14  
 21696  
 hypothetical protein F33C8.3 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 17-Mar-2000  
 C:Accession: T21696  
 R:percy, C.  
 submitted to the EMBL Data Library, February 1996  
 A:Reference number: Z19460  
 A:Accession: T21696  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-282 <W1>  
 A:Cross-references: EMBL:Z69790; PIDN:CAA93655.1; GSPDB:GN00028; CESP:F33C8.3  
 A:Experimental source: clone F33C8  
 C:Genetics:  
 A:Gene: CESP:F33C8.3  
 A:Map position: X  
 A:Interons: 20/3; 51/1; 88/3; 146/3; 197/1; 235/3  
 C:Superfamily: CD9 antigen

Query Match 21.7%; Score 255.5; DB 2; Length 282;  
 Best Local Similarity 24.9%; Pred. No. 2.3e-15;  
 Matches 70; Conservative 45; Mismatches 99; Indels 67; Gaps 9;

1 KYLLFVNFVFMALAGVILGVALMLRHPDPTSLYLTLTILNSLCPSGNILTPIL 59  
 9 RIVFLPFLFAMLSGVVFLGIVLWLPDPAASDPAL--HSTPGAFRYVG-WFLVGAG 64

60 AVGVAMFVFGICGALIOESQCLGTFPTCLVILPACVAAAGVFNKQIADVQK 118  
 65 AIIIVGVFGICGAMKNOQAL-AFPCIIILAFPLEIAAVTLFHNQENH-----KHVVE 119

119 QALQQAQVMDANNAAKAVVTFHETLNCSSNALTTITL 158  
 120 SSMYDTJTRNRYSSBFAKDFDYOEKRECCGVVTTYDMLSARMDABSTQLRYNBDAG 179

159 -----ILNSLCPSGNILTPIL-----LLQDCHQ 182  
 180 RIEHGIGAFGNKGTGGRVPSCCNHEKLSYNNCGRSQAPLTVTAQFINTGCA 239

183 KIDELFSKLYLIGIAIIVAVVIMIFEMILSMVLCCGIRNS 223  
 240 AVESVSSLSLIYGVGVLCIVQLGIVLSMTLCCCKGNS 280

Db

RESULT 15  
 A59265  
 tetraspan TSPAN-4 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 02-Jun-2000  
 C:Accession: A59265  
 R:Rod, S.C.; Doctor, V.S.; Levy, S.  
 Biochim. Biophys. Acta 1399, 101-104, 1998  
 A:Title: Sequences and expression of six new members of the tetraspanin/TM4SF family.  
 A:Reference number: A59258; MUID:98390278; PMID:9714763  
 A:Accession: A59265  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 1-238 <TOD>  
 A:Cross-references: GB:AF054841; NID:92997746; PIDN:AAC69717.1; PID:92997747  
 C:Genetics:  
 A:Gene: TSPAN-4  
 C:Superfamily: CD9 antigen

Query Match 21.6%; Score 254.5; DB 2; Length 238;  
 Best Local Similarity 27.8%; Pred. No. 2.4e-15;  
 Matches 66; Conservative 43; Mismatches 107; Indels 21; Gaps 5;

1 KYLLFVNFVFMALAGVILGVALMLRHPDPTSLYLTLTILNSLCPSGNILTPIL 59

```
ch completed: December 9, 2003, 10:32:50
time : 38 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

rotein - protein search, using sw model

on: December 9, 2003, 08:30:37 ; Search time 23 Seconds  
(without alignments)

462.088 Million cell updates/sec

ect score: US-10-035-914-2

ence: 1 KYLFFVNFVFLAGVILG.....IFEMILSMVLCGGIRNSSVY 226

ing table: BLOSUM62

Gapop 10.0 , Gapext 0.5

ched: 127863 seqs, 47026705 residues

1 number of hits satisfying chosen parameters: 127863

mm DB seq length: 0

mm DB seq length: 200000000

-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

base : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

lc	Score	Query Match	Length	DB	ID	Description
1	1177	99.7	236	1	CD81_MOUSE	P35762 mus musculu
2	1120	94.9	236	1	CD81_RAT	O62745 rattus norv
3	1104	93.6	236	1	CD81_CERAE	O97703 cercoptihc
4	1103	93.5	236	1	CD81_HUMAN	P18582 homo sapien
5	1100	93.2	236	1	CD81_SAGOE	O90095 saaguins oe
6	528.5	44.8	225	1	CD9_BOVIN	P30933 bos taurus
7	525.5	44.5	225	1	CD9_FITG	O8Wmq3 sus scrofa
8	522.5	44.3	225	1	CD9_RAT	P40241 rattus norv
9	518.5	43.9	227	1	CD9_HUMAN	P41926 homo sapien
10	515.5	43.7	225	1	CD9_FELCA	P40239 felis silve
11	512.5	43.4	227	1	CD9_CERAE	P30409 cercoptihc
12	510.5	43.3	225	1	CD9_MOUSE	P40240 mus musculu
13	462.5	39.2	221	1	TSN2_RAT	O9J1v1 rattus norv
14	458.5	38.9	221	1	TSN2_MOUSE	O92236 mus musculu
15	453.5	38.4	221	1	TSN2_HUMAN	O60636 homo sapien
16	298	25.3	237	1	T4S3_HUMAN	P19075 homo sapien
17	297.5	25.2	218	1	CD53_MOUSE	O61451 mus musculu
18	286.5	24.3	218	1	CD53_RAT	P24485 rattus norv
19	272.5	23.1	239	1	TNE5_HUMAN	O75954 homo sapien
20	272.5	23.1	267	1	CD82_HUMAN	P27701 homo sapien
21	265.5	22.5	219	1	CD83_HUMAN	P19397 homo sapien
22	262	22.2	266	1	CD82_MOUSE	P40237 mus musculu
23	254.5	21.6	238	1	T4S7_HUMAN	O14817 homo sapien
24	252.5	21.4	253	1	T4S7_MOUSE	O15566 mus musculu
25	250	21.2	256	1	CD82_RAT	O70352 rattus norv
26	248.5	21.1	253	1	CD51_RAT	O9qz26 rattus norv
27	247.5	21.0	238	1	T4S7_MOUSE	O9dck3 mus musculu
28	244.5	20.7	253	1	CD51_HUMAN	P4509 homo sapien
29	241.5	20.5	253	1	CD51_CERAE	O9wmy2 cercoptihc
30	220.5	18.7	241	1	TSN1_HUMAN	O60635 homo sapien
31	209	17.7	249	1	T4S2_HUMAN	P41732 homo sapien
32	207	17.5	237	1	CD63_MOUSE	P41731 mus musculu
33	206	17.5	268	1	T4S9_HUMAN	O60628 homo sapien

34	205	17.4	245	1	T4S6_HUMAN	O43657 homo sapien
35	205	17.4	249	1	T4S2_MOUSE	O62283 mus musculu
36	201	17.0	253	1	T4S8_MOUSE	O9gy33 mus musculu
37	199	16.9	253	1	T4S8_HUMAN	O60637 homo sapien
38	197.5	16.7	236	1	CD63_BOVIN	O9xk22 bos taurus
39	196	16.6	245	1	T4S6_MOUSE	O70401 mus musculu
40	195	16.5	237	1	CD63_RAT	P28648 rattus norv
41	194.5	16.5	281	1	CD37_HUMAN	P11049 homo sapien
42	187.5	15.9	281	1	CD37_MOUSE	O61470 mus musculu
43	187	15.8	237	1	CD63_RABIT	O28709 oryctolagus
44	180.5	15.3	281	1	CD37_RAT	P31053 rattus norv
45	175	14.8	305	1	TNE2_HUMAN	O95859 homo sapien

## ALIGNMENTS

RESULT 1	CD81_MOUSE	STANDARD;	PRT;	236 AA.
AC	P35762;			
DT	01-JUN-1994 (Rel. 29, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	CD81 antigen (26 kDa cell surface protein TAPA-1) (target of the			
DE	antiproliferative antibody 1).			
GN	CD81 OR TAPAL.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
NCBI_TaxID=10090;				
[1]				
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91318144; PubMed=1650385;			
RA	Andria M.L., Hsieh C.L., Oren R., Francke U., Levy S.;			
RT	"Genomic organization and chromosomal localization of the TAPA-1			
RT	gene.";			
RL	J. Immunol. 147:1030-1036(1991).			
RN	[2]			
RP	SEQUENCE OF 11-236 FROM N.A.			
RC	TISSUE=Heart;			
RA	Duff K., Parsons J.;			
RL	Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.			
CC	-1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE REGULATION OF LYMPHOMA			
CC	CELL GROWTH.			
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.			
CC	-1- PTM: NOT GLYCOSYLATED (PROBABLY).			
CC	-1- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL; S45012; AAB19417.1; -			
DR	EMBL; S44957; AAB19417.1; JOINED.			
DR	EMBL; S44966; AAB19417.1; JOINED.			
DR	EMBL; S45001; AAB19417.1; JOINED.			
DR	EMBL; S45008; AAB19417.1; JOINED.			
DR	EMBL; S45010; AAB19417.1; JOINED.			
DR	EMBL; X59047; -; NOT_ANNOTATED_CDS.			
DR	PIR; A46472; A46472.			
DR	MGI; MGI:1096398; Cdb1.			
DR	InterPro: IPR000301; Transmem_4.			
DR	Pfam: PF00335; Transmembrane4; 1.			
DR	PRINTS; PR00259; TMFOUR.			
DR	PROSITE; PS00421; TM4_1; 1.			
KW	Transmembrane; Antigen.			
FT	DOMAIN 1 33			
FT	TRANSMEM 13 33			
				CYTOPLASMIC (POTENTIAL).
				POTENTIAL.

DOMAIN 34 57 EXTRACELLULAR (POTENTIAL).  
TRANSMEM 58 78 POTENTIAL.  
DOMAIN 79 89 CYTOPLASMIC (POTENTIAL).  
TRANSMEM 90 115 POTENTIAL.  
TRANSMEM 116 201 EXTRACELLULAR (POTENTIAL).  
TRANSMEM 202 226 POTENTIAL.  
DOMAIN 227 236 CYTOPLASMIC (POTENTIAL).  
DOMAIN 173 173 T -> S (IN REF. 2).  
SEQUENCE 236 AA; 25828 MW; 7FCB386F79FA49 CRC64;

CY Match 99.7%; Score 1177; DB 1; Length 236;  
- Local Similarity 99.6%; Pred. No. 2,4e-93;  
- Cons 225; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 KYLFFNFVFWMLAGVILGVALMLRHPDPTSLYLKLNKPAKPTFYGYIILAVGA 60  
11 KYLFFNFVFWMLAGVILGVALMLRHPDPTSLYLKLNKPAKPTFYGYIILAVGA 70  
61 VMWVFGFCYGAIOESQCLGTFPTCLVILPACVAVAGIWFVNDQIAKVKQFYDOA 120  
71 VMWVFGFCYGAIOESQCLGTFPTCLVILPACVAVAGIWFVNDQIAKVKQFYDOA 130  
121 LQAAVMDDDANNAKAVVTFHETLNCSSNALTTTLTILNLSLCPSSGNILTPLLQDDC 180  
131 LQAAVMDDDANNAKAVVTFHETLNCSSNALTTTLTILNLSLCPSSGNILTPLLQDDC 190  
181 HOKIDELFSGKLYLIGIAIVAVIMIFEMILSMWLCGGIRNSSY 226  
191 HOKIDELFSGKLYLIGIAIVAVIMIFEMILSMWLCGGIRNSSY 236

1 2  
181 RAT STANDARD; PRT; 236 AA.  
16-OCT-2001 (Rel. 40, Created)  
16-OCT-2001 (Rel. 40, Last sequence update)  
16-OCT-2001 (Rel. 40, Last annotation update)  
181 antigen (26 kDa cell surface protein TAPA-1) (Target of the antiproliferative antibody 1).  
181 OR TAPA-1.  
Rattus norvegicus (Rat).  
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
[1] JCB1\_TaxID=10116;

SEQUENCE FROM N.A.  
STRAIN=Sprague-Dawley; TISSUE=Brain cortex;  
MEDLINE=96346153; PubMed=8757260;  
Reisert B.E., Yang L., Irwin M.H.;  
Astrocyte growth, reactivity, and the target of the antiproliferative antibody, TAPA-1;  
Neurosci. 16:5478-5487(1996).

1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE REGULATION OF LYMPHOMA CELL GROWTH.

1- SUBCELLULAR LOCATION: Integral membrane protein.  
1- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.

-----  
this SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

MBL; 019894; AAC3103.1; -  
InterPro; IPR000301; Transmem 4.  
fam; PF00335; transmembrane4; 1.  
PRINTS; PR00259; TMPOUR.  
PROSITE; PS00421; TM4.1; 1.  
Transmembrane; Antigen.  
DOMAIN 1 12  
CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 13 33 POTENTIAL.  
FT DOMAIN 34 57 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 58 78 POTENTIAL.  
FT DOMAIN 79 89 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 90 115 POTENTIAL.  
FT DOMAIN 116 201 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 202 226 POTENTIAL.  
FT DOMAIN 227 236 CYTOPLASMIC (POTENTIAL).  
SEQUENCE 236 AA; 25888 MW; DCC48F38EB19BDP5 CRC64;

Query Match 94.9%; Score 1120; DB 1; Length 236;  
Best Local Similarity 94.2%; Pred. No. 1.7e-88;  
Matches 213; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

1 KYLFFNFVFWMLAGVILGVALMLRHPDPTSLYLKLNKPAKPTFYGYIILAVGA 60  
11 KYLFFNFVFWMLAGVILGVALMLRHPDPTSLYLKLNKPAKPTFYGYIILAVGA 70  
61 VMWVFGFCYGAIOESQCLGTFPTCLVILPACVAVAGIWFVNDQIAKVKQFYDOA 120  
71 VMWVFGFCYGAIOESQCLGTFPTCLVILPACVAVAGIWFVNDQIAKVKQFYDOA 130  
121 LQAAVMDDDANNAKAVVTFHETLNCSSNALTTTLTILNLSLCPSSGNILTPLLQDDC 180  
131 LQAAVMDDDANNAKAVVTFHETLNCSSNALTTTLTILNLSLCPSSGNILTPLLQDDC 190  
181 HOKIDELFSGKLYLIGIAIVAVIMIFEMILSMWLCGGIRNSSY 226  
191 HOKIDELFSGKLYLIGIAIVAVIMIFEMILSMWLCGGIRNSSY 236

## RESULT 3

CD81\_CERAB STANDARD; PRT; 236 AA.  
16-OCT-2001 (Rel. 40, Created)  
16-OCT-2001 (Rel. 40, Last sequence update)  
16-OCT-2001 (Rel. 40, Last annotation update)  
CD81 antigen.  
GN  
CD81.  
Cercopithecus aethiops (Green monkey) (Griivet).  
OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
OC Cercopitheidae; Cercopitheidae.  
OX NCBI\_TaxID=9534;

SEQUENCE FROM N.A.  
RA Levy S., Kuo C.C.;  
RT "African green monkey CD81 cDNA";  
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.  
CC 1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE REGULATION OF LYMPHOMA CELL GROWTH.

CC 1- SUBCELLULAR LOCATION: Integral membrane protein.  
CC 1- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

DR EMBL; AF116599; AAD11439.1; -  
DR InterPro; IPR000301; Transmem 4.  
DR Pfam; PF00335; transmembrane4; 1.  
DR PRINTS; PR00259; TMPOUR.  
DR PROSITE; PS00421; TM4.1; 1.  
KW Transmembrane; Antigen.  
FT TRANSMEM 1 12  
FT DOMAIN 13 33  
FT TRANSMEM 34 57  
FT TRANSMEM 58 78  
CYTOPLASMIC (POTENTIAL).  
EXTRACELLULAR (POTENTIAL).  
POTENTIAL.

DOMAIN 79 89 CYTOPLASMIC (POTENTIAL).  
 TRANSMEM 90 115 POTENTIAL.  
 DOMAIN 116 201 EXTRACELLULAR (POTENTIAL).  
 TRANSMEM 202 226 POTENTIAL.  
 DOMAIN 227 236 CYTOPLASMIC (POTENTIAL).  
 SEQUENCE 236 AA; 25758 MW; E07A8BEC27637A4A CRC64;  
 ery Match 93.6%; Score 1104; DB 1; Length 236;  
 at Local Similarity 92.0%; Pred. No. 4e-87;  
 tches 208; Conservative 13; Mismatches 5; Indels 0; Gaps 0;

1 KTLFPNFPNFMVLGGVILGVALLMHPDQTTSLYLEGKRAKPAFTFYGIYLLAVGA 60  
 |||||  
 11 KTLFPNFPNFMVLGGVILGVALLMHPDQTTSLYLEGKRAKPAFTFYGIYLLAVGA 70  
 |||||  
 61 VMFVGFAGCYGAIQBSQCLIGFTFCLVILFACVAAAGIWMGVNDQIAKVKQFYDQA 120  
 |||||  
 71 VMFVGFAGCYGAIQBSQCLIGFTFCLVILFACVAAAGIWMGVNDQIAKVKQFYDQA 130  
 |||||  
 121 LQAAVMDDDANNAKAVVKTFFHETLNCQGSNALTTLTTLIRNSLCPSGGNILTPLLQDPC 180  
 |||||  
 131 LQAAVMDDDANNAKAVVKTFFHETLNCQGSNALTTLTTLIRNSLCPSGGNILTPLLQDPC 190  
 |||||  
 181 HOKIDELPSGKLYLIGIAIVAVVIMFPMILSMVLCGIRNSSYV 226  
 |||||  
 191 HOKIDELPSGKLYLIGIAIVAVVIMFPMILSMVLCGIRNSSYV 236  
 |||||

LT 4  
 HUMAN  
 CD81 HUMAN STANDARD; PRT; 236 AA.  
 P18582;  
 01-NOV-1990 (Rel. 16, Created)  
 01-NOV-1990 (Rel. 16, Last sequence update)  
 15-SEP-2003 (Rel. 42, Last annotation update)  
 CD81 antigen (26 kDa cell surface protein TAP-1) (Target of the  
 antiproliferative antibody 1).  
 CD81 OR TAP-1.  
 Homo sapiens (Human), and  
 Pan troglodytes (Chimpanzee).  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 NCBI\_TaxID=9606, 9598;  
 [1]  
 SEQUENCE FROM N.A.  
 SPECIES=Human;  
 MEDLINE=90318365; PubMed=1695320;  
 Oren R., Takahashi S., Doss C., Levy R., Levy S.,  
 "TAP-1, the target of an antiproliferative antibody, defines a new  
 family of transmembrane proteins.";  
 Mol. Cell. Biol. 10:4007-4015(1990).  
 [2]  
 SEQUENCE FROM N.A.  
 SPECIES=Human; TISSUE=Brain;  
 MEDLINE=22388257; PubMed=12477932;  
 Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Datchenko L., Maruina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stepien M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 Brownstein M.J., Ueda T.B., Toshiyuki S., Carrinci P., Prange C.,  
 Raba S.S., Loguelfano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,  
 Boak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gamarale P.H.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Faley J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 Whitting R.M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M.,  
 Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
 Schnerch A., Schein J.E., Jones S.J.M., Matra M.A.,  
 "Generation and initial analysis of more than 15,000 full-length

RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=P.troglodytes;  
 RA Levy S., Kuo C.C.;  
 RT "CD81 cDNA sequence derived from chimpanzee cells.";  
 RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP TOPOLOGY.  
 RX MEDLINE=91317825; PubMed=1860863;  
 RA Levy S., Nguyen V.O., Andria M.L., Takahashi S.;  
 RT "Structure and membrane topology of TAP-1.";  
 RL J. Biol. Chem. 266:14597-14602(1991).  
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE REGULATION OF LYMPHOMA  
 CC CELL GROWTH. INTERACTS WITH A 16-KDA LBU-13 PROTEIN TO FORM A  
 CC COMPLEX POSSIBLY INVOLVED IN SIGNAL TRANSDUCTION. MAY ACTS A THE  
 CC VIRAL RECEPTOR FOR HCV.  
 CC -1- SUBUNIT: Interacts with Hepatitis C virus (HCV) glycoprotein E2.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: HEPATOLYMPHOID, NEUROECTODERMAL, & MESENCHYMAL  
 CC TUMOR CELL LINES.  
 CC -1- PTM: NOT GLYCOSYLATED (PROBABLE).  
 CC -1- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.  
 CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD81 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd81.htm".

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

DR EMBL; M33680; AAA36653.1; -  
 DR EMBL; BC002978; AAH02978.1; -  
 DR EMBL; AF116600; AA01440.1; -  
 DR PIR; A35649; A35649.  
 DR PDB; 1G8Q; 21-FEB-01.  
 DR PDB; 1IV5; 28-JAN-03.  
 DR Genew; HGNC:1701; CD81.  
 DR MIM; 186845; -  
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.  
 DR GO; GO:0008283; P:cell proliferation; TAS.  
 DR GO; GO:0006952; P:defense response; TAS.  
 DR InterPro; IPR00301; Transmem 4.  
 DR Pfam; PF00335; transmembrane4; 1.  
 DR PRINTS; PR00259; TMFOUR.  
 DR PROSITE; PS00421; TM4\_1; 1.  
 KW Transmembrane; Antigen; 3D-structure.  
 FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 13 33 POTENTIAL.  
 FT DOMAIN 34 63 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 64 84 POTENTIAL.  
 FT DOMAIN 85 89 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 90 112 POTENTIAL.  
 FT DOMAIN 113 201 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 202 224 POTENTIAL.  
 FT DOMAIN 225 236 CYTOPLASMIC (POTENTIAL).  
 SEQUENCE 236 AA; 25809 MW; B99BD761AC918A CRC64;

Query Match 93.5%; Score 1103; DB 1; Length 236;  
 Best Local Similarity 91.6%; Pred. No. 4.9e-87;  
 Matches 207; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

OY 1 KTLFPNFPNFMVLGGVILGVALLMHPDQTTSLYLEGKRAKPAFTFYGIYLLAVGA 60  
 |||||  
 DB 11 KTLFPNFPNFMVLGGVILGVALLMHPDQTTSLYLEGKRAKPAFTFYGIYLLAVGA 70  
 |||||  
 OY 61 VMFVGFAGCYGAIQBSQCLIGFTFCLVILFACVAAAGIWMGVNDQIAKVKQFYDQA 120  
 |||||  
 DB 71 VMFVGFAGCYGAIQBSQCLIGFTFCLVILFACVAAAGIWMGVNDQIAKVKQFYDQA 130  
 |||||

121 LQAVMDDANNAAVVKTFHETLNCSSNALTTTLTILNSLCPSGGNLTPLQDDC 180  
 131 LQAVVDDANNAAVVKTFHETLNCSSSTLTALTSTVAKNLCSSGSSITSNLFKEDC 190  
 181 HOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGGINSSSV 226  
 191 HOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGGINSSSV 236

T 5

CD81 SAGOR STANDARD; PRT; 236 AA.  
 29NOJ3;  
 28-FEB-2003 (Rel. 41, Created)  
 28-FEB-2003 (Rel. 41, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 CD81 protein.

Saguinus oedipus (Cotton-top tamarin).  
 Sakyota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Saguinus.  
 NCBI\_TaxID=9490;  
 (1)  
 3SEQUENCE FROM N.A.  
 TISSUE=Liver;  
 MEDLINE=20304985; PubMed=10846074;  
 Meola A., Bardelaci A., Bruni E.B., Cerretani M., Pezzanera M.,  
 Cecacci A., Vitelli A., Levy S., Nicotia A., Traboni C.,  
 Cecacci A., Scarselli E.,  
 Binding of hepatitis C virus E2 glycoprotein to CD81 does not  
 correlate with species permissiveness to infection.\*  
 J. Virol. 74:5933-5938 (2000).

CELL GROWTH (BY SIMILARITY).  
 - FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE REGULATION OF LYMPHOMA  
 CELL GROWTH.  
 - SUBCELLULAR LOCATION: Integral membrane protein.  
 - SIMILARITY: Belongs to the tetraspanin (TM4SF) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: AJ250197; CAB89875.1; -  
 InterPro: IPR000301; Transmem\_4.  
 Pfam: PF00335; transmembrane4; 1.  
 PRINTS: PR00259; TMFOUR.  
 PROSITE: PS00421; TM4\_1; 1.  
 Transmembrane; Antigen.

CD81 SAGOR STANDARD; PRT; 236 AA.  
 29NOJ3;  
 28-FEB-2003 (Rel. 41, Created)  
 28-FEB-2003 (Rel. 41, Last sequence update)  
 28-FEB-2003 (Rel. 41, Last annotation update)  
 CD81 protein.

Y Match 93.2%; Score 1100; DB 1; Length 236;  
 Local Similarity 91.6%; Pred. No. 8.8e-87;  
 Conservative 14; Mismatches 5; Indels 0; Gaps 0;

1 KYLFVFNFFVMTAGVILGVALMLRHPDPTSLTLYELGNKPAENTFYGIYLIANGA 60  
 11 KYLFVFNFFVMTAGVILGVALMLRHPDPTSLTLYELGNKPAENTFYGIYLIANGA 70  
 61 VMMFVGLCYCAIGESQCLGTFTCTVILFACGVAAGMGFVKDIOIADVKQFYDOA 120

Db 71 VMMFVGLCYCAIGESQCLGTFTCTVILFACGVAAGMGFVKDIOIADVKQFYDOA 130  
 Qy 121 LQAVMDDANNAAVVKTFHETLNCSSNALTTTLTILNSLCPSGGNLTPLQDDC 180  
 Db 131 LQAVVDDANNAAVVKTFHETLNCSSSTLTALTSTVAKNLCSSGSSITSNLFKEDC 190  
 Qy 181 HOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGGINSSSV 226  
 Db 191 HOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGGINSSSV 236

RESULT 6

CD9 BOVIN  
 ID CD9 BOVIN STANDARD; PRT; 225 AA.  
 AC P30532;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE CD9 antigen.

OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;  
 OC Bovidae; Bovinae; Bos.  
 OC NCBI\_TaxID=9913;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Ocular ciliary epithelium;  
 RA MEDLINE=93054422; PubMed=1339429;  
 RX Martin-Alonso J.M., Hernandez N., Ghosh S., Coca-Prados M.,  
 "Molecular cloning of the bovine CD9 antigen from ocular ciliary  
 epithelial cells.\*"  
 RT J. Biochem. 112:63-67 (1992).

CC - FUNCTION: INVOLVED IN PLATELET ACTIVATION AND AGGREGATION.  
 CC - SUBCELLULAR LOCATION: Integral membrane protein.  
 CC - SIMILARITY: Belongs to the tetraspanin (TM4SF) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: M81720; AAA30439.1; -  
 PIR: JX0221; JX0221.  
 InterPro: IPR000301; Transmem\_4.  
 Pfam: PF00335; transmembrane4; 1.  
 PRINTS: PR00259; TMFOUR.  
 PROSITE: PS00421; TM4\_1; 1.  
 Glycoprotein; Antigen; Transmembrane.

CD9 BOVIN STANDARD; PRT; 225 AA.  
 AC P30532;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE CD9 antigen.

Query Match 44.8%; Score 528.5; DB 1; Length 225;  
 Best Local Similarity 46.2%; Pred. No. 4.1e-38;  
 Matches 103; Conservative 43; Mismatches 66; Indels 11; Gaps 5;

1 KYLFVFNFFVMTAGVILGVALMLRHPDPTSLTLYELGNKPAENTFYGIYLIANGA 60  
 10 KYLFVFNFFVMTAGVILGVALMLRHPDPTSLTLYELGNKPAENTFYGIYLIANGA 65



```

121 100A VMD DANNAA VVKT FHE RT LNC CGS NAL T L T T L I N S I C P S G S N I L T P L Q D C 180
126 Y N L K N N D B Q - R E T I K A H I A L D C G - - L T S V P Q F L T D - C P - K A L I D S L K T R P C 178
181 H K I D B L F S G K Y L I G I A I V A V A T M I F E M I L S K V L C C G I R S 223
179 P R A I D E I F R S K F H I L G A V G I A N V M I R G A V P S M I L C A I R N 221

```

LT 7	CD9 FIG	STANDARD;	PRT;	225 AA.
FIG	GBM03;			
	28-FEB-2003	(Rel. 41, Created)		
	28-FEB-2003	(Rel. 41, Last sequence update)		
	28-FEB-2003	(Rel. 41, Last annotation update)		
	CD9 antigen.			
	CD9.			

*Sus scrofa* (Pig).  
Bakaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
NCBI\_TaxID=9823;

SEQUENCE FROM N.A.  
Yubero N., Barbanchó M.J., Llanes D., Garrido J.J.:  
"Molecular cloning of the pig homolog of tetraspanin CD9 antigen."  
Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.  
FUNCTION: INVOLVED IN PLATELET ACTIVATION AND ADHESION (2)

- 1- SUBCELLULAR LOCATION: Integral membrane protein.
- 1- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sb.ch/announce/> to send an email to [license@isb-sb.ch](mailto:license@isb-sb.ch)).

```
EMBL; AY072785; AAL68966.1; -.
InterPro; IPR000301; Transmem_4.
Pfam; PF00335; transmembrane4; 1.
PRINTS; PR00259; TMFOUR.
PROSITE; PS00421; TM4_1; 1.
```

INIT MET	GLYCOPROTEIN; ANTIGEN;	TRANSMEMBRANE; LIPOPROTEIN. BY SIMILARITY.
0	0	CYTOPLASMIC (POTENTIAL).
1	11	POTENTIAL.
12	32	EXTRACELLULAR (POTENTIAL).
33	52	POTENTIAL.
53	73	CYTOPLASMIC (POTENTIAL).
74	84	POTENTIAL.
85	108	EXTRACELLULAR (POTENTIAL).
109	192	POTENTIAL.
193	218	CYTOPLASMIC (POTENTIAL).
219	225	N-LINKED (GLCNAC. . .) (POTENTIAL).
49	50	N-LINKED (GLCNAC. . .) (POTENTIAL).
50	50	FF80FB39BC11545 CBC64;
225 AA;	25070 MM;	

try Match	44.5%	Score 525.5;	DB 1;	length 225;
Local Similarity	47.1%;	Pred. No. 7.3e-38;		
ches 105; Conservative	37;	Mismatches 70;	Indels 11;	Gaps 5

```

1 KLLLEFVENVFMLAGVILGVALMLRHPDPTTSLLEYLEGNKPAENPFYGIYILAVGA 60
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
10 KLLLGGFNFIFMLAGIAVLAIGLMLRFDSTQYSIFQENNN---SSEFYGVILIGAGA 65
  ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
61 VMVFVGLGCTAIGESQCLLGFFFLCVILFACBVAAGVWGTVNDDQIAKDVYKPTDQA 120

```

Db 66 LMMVGFGLCCGAVDSQCMGLFPGFLVFLFALEIMAAIMGSHKQVIXKEYODFPRDT 125  
 Qy 121 LQAAVNDDDANNAKAAVKTFFHTLMCCSNALVTTLTILNSICPSSGNITPLQODC 180  
 Db 126 YNLTAKKEDPQ--RETLLAIHYALDCCG---LMGFEVRLDLD-ICPQ-RDVLSSLPMKPC 178  
 Qy 181 HOKIDELFSGKLYLIGIAIAVAIVAMIFEMILSNVLCGGINRS 223  
 Db 179 PEAIKREVPQMKFHTIIGAVGIGIAVAMITGMFPMILCAIIRS 221

RESULT 8		
CD9_RAT		
ID	CD9_RAT	STANDARD; PRT; 225 AA.
AC	P40241;	
DT	01-FEB-1995	(Rel. 31, Created)
DT	01-FEB-1995	(Rel. 31, Last sequence update)
DT	16-OCT-2001	(Rel. 40, Last annotation update)
DE	cd9 antigen.	

SA  
CO5.  
OC Rattus norvegicus (Rat).  
OC Eulayrola; Metascia; Chordata; Craniata; Vertebrata; Eumleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus  
OX NCBI\_TaxID=10116;  
rel

RA SEQUENCE FROM N.A.  
RE  
RT STRAIN=Sprague-Dawley; TISSUE=sciatic nerve;  
RC STRAIN=Sprague-Dawley; PubMed:7833164;  
RX MEDLINE=95123481; PubMed:7833164;  
RA Kaprielian Z., Cho K.O., Haffajiryou M., Patterson P.H.;  
RT "CD9, a major platelet cell surface glycoprotein,  
is expressed in the nervous system."; A ROCA antigen  
LT J. Neurosci. 15:562-573(1995).

CC	-1-	FUNCTION: INVOLVED IN PLATELET ACTIVATION AND AGGREGATION
CC	-1-	(BY SIMILARITY).
CC	-1-	SUBCELLULAR LOCATION: Integral membrane protein.
CC	-1-	TISSUE SPECIFICITY: EXPRESSED IN THE PERIPHERAL NERVOUS SYSTEM
CC	-1-	SIMILARITY: Belongs to the tetraepanin (TM4SF) family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/isb-sib.ch>).

[illegible]

```

Query Match      44.3%; Score 522.5; DB 1; Length 225;
Best Local Similarity 45.1%; Pred. No. 1.3e-37;
Matches 101; Conservative 39; Mismatches 71; Indels 13; Gaps 5

QY      1 KLLLEVFVFWMLAGVILGVALMLRHDPTTSLILYLELGNKKEAPNTFYVGIYLLAVG 60
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

10 KYLLFGNFIFMLAGIIVLALGIMLRPDSOTKS1PEQBNH-----SSFTYGVILLIGAGA 65  
 11 VMFVFGYGCYGAIOESQCLIGTFTCLVILFACEVAGIWMGFVNKOIKNDVQVYDQA 120  
 12 LMLVGLFGCCGAGVQESQCMGLFFGLVILFALFIAAAWGYHNDVKEIKELQEFYKDT 125  
 131 LQQAAMDNDANANAAVVKTFHETLNCCG-SNALTTTLTTLTTLNLSLCPGSGNLIPLPLQDO 179  
 126 YOKLRNDEPQ--RETLLKALHMLNCCGAGVGVQFISDI-----CK-KQVLESPOVKS 177  
 180 CHQKIDELFSGKYLIGIILVAIVAVIPEMILSMVLCGGRNS 223  
 178 CPDAIDEVFHSKPHIIGAVGIGIIVMIFGMIFSMILCAIRRS 221

T-9  
 UMAN  
 CD9 HUMAN STANDARD; PRT; 227 AA.  
 P21926; J96884;  
 01-MAY-1991 (Rel. 18, Created)  
 01-APR-1993 (Rel. 25, Last sequence update)  
 15-SEP-2003 (Rel. 42, Last annotation update)  
 CD9 antigen (P24) (Leukocyte antigen M1C3) (Motility-related protein) (MRP-1).  
 CD9 OR M1C3.  
 Homo sapiens (Human).  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NCBI\_TaxID=9606;  
 (1)  
 SEQUENCE FROM N.A., AND SEQUENCE OF 1-4.  
 MEDLINE=91093112; PubMed=1840589;  
 Boucheix C., Benoit P., Prachet P., Billard M., Worthington R.B., Zagnon J., Uzan G.;  
 "Molecular cloning of the CD9 antigen. A new family of cell surface proteins.";  
 J. Biol. Chem. 266:117-122(1991).  
 (2)  
 SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 MEDLINE=91244846; PubMed=2037603;  
 Janza F., Wolf D., Fox C.F., Kleffer N., Seyer J.M., Fried V.A., Goughlin S.R., Phillips D.R., Jennings L.K.;  
 "cDNA cloning and expression of platelet P24/CD9. Evidence for a new family of multiple membrane-spanning proteins.";  
 J. Biol. Chem. 266:10638-10645(1991).  
 (3)  
 SEQUENCE FROM N.A.  
 MEDLINE=92078843; PubMed=1720807;  
 Miyake M., Koyama M., Sano M., Ikegama S.;  
 "Identification of the motility-related protein (MRP-1), recognized by monoclonal antibody M31-15 which inhibits cell motility.";  
 J. Exp. Med. 174:1347-1354(1991).  
 (4)  
 SEQUENCE FROM N.A.  
 TISSUE=Leukocyte;  
 MEDLINE=93252369; PubMed=8486348;  
 Rubinstejn B., Benoit P., Billard M., Plaisance S., Prenant M., Uzan G., Boucheix C.;  
 "Organization of the human CD9 gene.";  
 Genomics 16:132-138(1993).  
 (5)  
 SEQUENCE FROM N.A.  
 TISSUE=Ovary;  
 MEDLINE=22388257; PubMed=12477932;  
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., White S.P., Zeeberg B., Buettow K.H., Scheafer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heisch F., Jatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Iatlenko M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.R., Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C., Jaba S.S., Lequellano N.A., Peters G.J., Abramson R.D., Millshy S.J., Josak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Rahay J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
 RA Schenker A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "regeneration and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [6]  
 RP SEQUENCE OF 1-20.  
 RC TISSUE=Platelet;  
 RA MEDLINE=90292223; PubMed=2358073;  
 RA Higashihara M., Takahata K., Yatom Y., Nakahara K., Kurokawa K.;  
 RT "Purification and partial characterization of CD9 antigen of human  
 RT platelets.";  
 RL FEBS Lett. 264:270-274(1990).  
 RN [7]  
 RP PHOSPHORYLATION.  
 RA MEDLINE=93327758; PubMed=7687539;  
 RA Chalupny N.J., Kanner S.B., Schieven G.L., Wee S., Gilliland L.K.,  
 RA Arafio A., Ledbetter J.A.;  
 RT "Tyrosine phosphorylation of CD19 in pre-B and mature B cells.";  
 RL EMBO J. 12:2691-2696(1993).  
 RN [8]  
 RP INTERACTION WITH PTGFRN.  
 RA MEDLINE=21216740; PubMed=11278880;  
 RA Charrin S., Le Naour F., Oualid M., Billard M., Faure G., Hanash S.M.,  
 RA Boucheix C., Rubinstein E.;  
 RT "The major CD9 and CD81 molecular partner. Identification and  
 RT characterization of the complexes.";  
 RL J. Biol. Chem. 276:14329-14337(2001).  
 CC -1- FUNCTION: INVOLVED IN PLATELET ACTIVATION AND AGGREGATION.  
 CC -1- SUBUNIT: Associates with CR2/CD21 and with PTGFRN/CD91.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED BY A VARIETY OF HEMATOPOIETIC  
 CC AND EPITHELIAL CELLS.  
 CC -1- PTM: PROTEIN EXISTS IN THREE FORMS WITH MOLECULAR MASSES BETWEEN  
 CC 22 AND 27 kDa, AND IS KNOWN TO CARRY COVALENTLY LINKED FATTY  
 CC ACIDS.  
 CC -1- PTM: PHOSPHORYLATED ON TYROSINE FOLLOWING B CELL ACTIVATION.  
 CC -1- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.  
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD9 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd9.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@db-eb.ch](mailto:license@db-eb.ch)).  
 CC -----  
 DR EMBL; S60489; AAC60586.1; JOINED.  
 DR EMBL; S60462; AAC60586.1; JOINED.  
 DR EMBL; S60463; AAC60586.1; JOINED.  
 DR EMBL; S60464; AAC60586.1; JOINED.  
 DR EMBL; S60700; AAC60586.1; JOINED.  
 DR EMBL; S60699; AAC60586.1; JOINED.  
 DR EMBL; S60465; AAC60586.1; JOINED.  
 DR EMBL; S60472; AAC60586.1; JOINED.  
 DR EMBL; M38690; AAA80320.1; -.  
 DR EMBL; L34068; AAA59982.1; -.  
 DR EMBL; L60111; CAA42708.1; -.  
 DR EMBL; L08118; -; NOT ANNOTATED CDS.  
 DR EMBL; L08120; AAA51955.1; ALT\_SEQ.  
 DR EMBL; L08121; AAA51956.1; -.  
 DR EMBL; L08122; AAA51957.1; -.  
 DR EMBL; L08123; AAA51958.1; -.  
 DR EMBL; L08124; AAA51959.1; -.

[illegible]

```

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; L35275; AAA92867.1; -.
DR EMBL; D30786; BAA06452.1; -.
DR InterPro; IPR000301; Transmem_4.
DR Pfam; PF00335; transmembrane4; 1.
DR PRINTS; PR00259; TM4P.
DR PROSITE; PS00421; TM4_1; 1.
KW Glycoprotein; Antigen; Transmembrane.
FT INIT MET 0 BY SIMILARITY.
FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 12 32 POTENTIAL.
FT DOMAIN 33 52 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 53 73 POTENTIAL.
FT DOMAIN 74 84 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 85 108 POTENTIAL.
FT DOMAIN 109 192 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 193 218 POTENTIAL.
FT DOMAIN 219 225 CYTOPLASMIC (POTENTIAL).
FT CONFLICT 82 S -> Y (IN REF. 2).
SQ SEQUENCE 225 AA; 24927 MW; 5A049774501B0383 CRC64;

Query Match 43.7%; Score 515.5; DB 1; Length 225;
Best Local Similarity 44.9%; Pred. No. 5,2e-37;
Matches 101; Conservative 42; Mismatches 67; Indels 15; Gaps 5;

QY 1 KYLLFVNFVFWLAVGLVALMLRHDPTQTSLLYLELGNKFAPIFTFYGIYLLAVGA 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 10 KYLLFGFNFVFWLAVGLVALMLRHDPTQTSIFPD-----SQPSFTYGIYLLAVGA 65
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 VMNVVGLFGCGAIGQESQCLGTFPTCLVLLPACVAAIGVFNVDQIAKDVKQFYDQA 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 66 LMLVGLFGCGAIGQESQCMGLFGFLVTFMIRIAAIGVSHKDEVAIQEVEFYKDT 125
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 LQGVVMDDDANNAKAVVVFTHETLNCQGSNAALTTLTTLIR--NSLCPSGNTLPFLQ 178
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 126 YNKKSKDDEP--RDTLKAITHVALDCG-----LNGVEQPISDICPQ-KDILSSITVX 176
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 179 DGHQIDELFSGLKYLIGIAIAIVAAVIMPEMLISVWLCCGIPNS 223
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 177 PCBAIKVEVFNKPHITIGAVGIAVVMIFGMLFPMILCAIARRS 221
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

RESULT 11	CD9_CERAE	STANDARD;	PRT;	227 AA.
ID	CD9_CERAE			
AC	P30403;			
DT	01-APR-1993	(Rel. 25, Created)		
DT	01-APR-1993	(Rel. 25, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	CD9 antigen (27 kDa diphtheria toxin receptor-associated protein) (DRAP27).			
GN	CD9.			
OS	Cercopithecus aethiops (Green monkey) (Grivel).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;			
OC	Cercopithecinae; Cercopithecus.			
OX	NCBI_TaxID=9534;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92394967; Pubmed=1523113;			
RA	Mitamura T., Iwamoto R., Umata T., Yomo T., Uraibe I., Tsuneoka M.,			
RT	"The 27-kD diphtheria toxin receptor-associated protein (DRAP27) from			

vero cells is the monkey homologue of human CD9 antigen: expression of DRAP27 elevates the number of diphtheria toxin receptors on toxin-sensitive cells. "

J. Cell Biol. 118:1389-1399(1992).

-1- FUNCTION: EXPRESSION OF DRAP27 ELEVATES THE NUMBER OF DIPHTHERIA TOXIN RECEPTORS ON TOXIN-SENSITIVE CELLS.

-1- SUBCELLULAR LOCATION: Integral membrane protein.

-1- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; D10726; BAA01569.1; -

PIR; A42929; A42929.

InterPro; IPR000301; Transmem 4.

Pfam; PF00335; Transmembrane4; 1.

PRINTS; PR00259; TMFOUR.

PROSITE; PS00421; TM4\_1; 1.

GLYCOPROTEIN; Antigen; Transmembrane.

INIT MET 0 BY SIMILARITY.

DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).

DOMAIN 12 32 EXTRACELLULAR (POTENTIAL).

DOMAIN 33 54 EXTRACELLULAR (POTENTIAL).

DOMAIN 55 75 CYTOPLASMIC (POTENTIAL).

DOMAIN 76 86 EXTRACELLULAR (POTENTIAL).

DOMAIN 87 110 EXTRACELLULAR (POTENTIAL).

DOMAIN 111 194 EXTRACELLULAR (POTENTIAL).

TRANSMEM 195 220 POTENTIAL.

DOMAIN 221 227 CYTOPLASMIC (POTENTIAL).

CARBOHYD 51 51 N-LINKED (GLCNAC. . .) (POTENTIAL).

CARBOHYD 52 52 N-LINKED (GLCNAC. . .) (POTENTIAL).

SEQUENCE 227 AA; 25300 MW; 88347487A590374A CRC64;

Query Match 43.4%; Score 512.5; DB 1; Length 227;

Best Local Similarity 44.4%; Pred. No. 9.4e-37;

Matches 100; Conservative 40; Mismatches 72; Indels 13; Gaps 5;

1 KYLLFVFNFMVLGAVILGVALMLRHPDPTSLYLKGNKPAFTFYGIYLLAVGA 60

10 KYLLFVFNFMVLGAVILGVALMLRHPDPTSLYLKGNKPAFTFYGIYLLAVGA 67

61 VMVVFGLGCTGAIQESQCLLGFPTCLVILFACGEVAGIWMGVNDDQIAKDYKQFYDQ 120

68 LMMLVGFLGCGAVQESQCLLGFPTCLVILFACGEVAGIWMGVNDDQIAKDYKQFYDQ 127

121 LQAVVDDDDANNAKAVVKTFFHTLNCSSNALTTTLTIIR--NSLCPSGNTLTPLLQD 178

128 YNKAKTKDQEQ--RETLKALHVALDCCG-----LACGVQSFISDICK-KDVALETTIK 178

179 DCHOKIDELFSKGLYLIGIAIVAVVIMFEMILSMVLCGGIRNS 223

179 SCDDPAIKVDFDNKFIIGAVGIGIIVMIFGMIFSMILCCAIRN 223

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

SEQUENCE FROM N.A.

RC STRAIN-BALB/C; TISSUE=Kidney;

RX MEDLINE=94054345; PubMed=8236164;

RA Rudinstein E., Billard M., Plaisance S., Prenant M., Boucheix C.;

RT "Molecular cloning of the mouse equivalent of CD9 antigen. "

Thromb. Res. 71:377-383(1993).

-1- FUNCTION: INVOLVED IN PLATELET ACTIVATION AND AGGREGATION

(BY SIMILARITY).

-1- SUBCELLULAR LOCATION: Integral membrane protein.

-1- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L08115; AAA37405.1; -

PIR; I49589; I49589.

MGI; MGI:88348; CD9.

InterPro; IPR000301; Transmem 4.

Pfam; PF00335; Transmembrane4; 1.

PRINTS; PR00259; TMFOUR.

PROSITE; PS00421; TM4\_1; 1.

GLYCOPROTEIN; Antigen; Transmembrane.

INIT MET 0 BY SIMILARITY.

DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).

DOMAIN 12 32 EXTRACELLULAR (POTENTIAL).

DOMAIN 33 52 EXTRACELLULAR (POTENTIAL).

DOMAIN 53 73 CYTOPLASMIC (POTENTIAL).

DOMAIN 74 84 EXTRACELLULAR (POTENTIAL).

DOMAIN 85 108 EXTRACELLULAR (POTENTIAL).

DOMAIN 109 192 EXTRACELLULAR (POTENTIAL).

TRANSMEM 193 218 POTENTIAL.

DOMAIN 219 225 CYTOPLASMIC (POTENTIAL).

CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).

SEQUENCE 225 AA; 25127 MW; 2BBB40B8D7C31BC0 CRC64;

Query Match 43.3%; Score 510.5; DB 1; Length 225;

Best Local Similarity 44.2%; Pred. No. 1.4e-36;

Matches 99; Conservative 41; Mismatches 71; Indels 13; Gaps 5;

1 KYLLFVFNFMVLGAVILGVALMLRHPDPTSLYLKGNKPAFTFYGIYLLAVGA 60

10 KYLLFVFNFMVLGAVILGVALMLRHPDPTSLYLKGNKPAFTFYGIYLLAVGA 65

61 VMVVFGLGCTGAIQESQCLLGFPTCLVILFACGEVAGIWMGVNDDQIAKDYKQFYDQ 120

66 LMMLVGFLGCGAVQESQCLLGFPTCLVILFACGEVAGIWMGVNDDQIAKDYKQFYDQ 125

121 LQAVVDDDDANNAKAVVKTFFHTLNCSSNALTTTLTIIRNSLCPSGNTLTPLLQD 179

126 YQKRSKQDQEQ--RETLKALHVALDCCG--SNTCPK-KQLLESQVMP 177

180 CHOKIDELFSKGLYLIGIAIVAVVIMFEMILSMVLCGGIRNS 223

178 CPEALISEVENNKHFIIGAVGIGIIVMIFGMIFSMILCCAIRS 221

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

CD9 antigen.

NCBI\_TaxID=10116;

[11]

SEQUENCE FROM N.A.

MEDLINE=20047448; PubMed=10582623;

Birling M.C., Tait S., Hardy R.J., Brophy P.J.;

"A novel rat tetraspanin protein in cells of the oligodendrocyte lineage.";

J. Neurochem. 73:2600-2608(1999).

-1- FUNCTION: May play a role in signalling in oligodendrocytes in the

early stages of their terminal differentiation into myelin-forming

glia and may also function in stabilizing the mature sheath.

-1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

-1- TISSUE SPECIFICITY: Expression is restricted to the nervous system.

-1- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL: AJ271442; CAB69827.1; -

InterPro: IPR000301; Transmem\_4.

Pfam: PF00335; transmembrane4; 1.

PRINTS: PR00259; TMFOUR.

PROSITE: PS00421; TM4.1; 1.

Glycoprotein; Transmembrane.

DOMAIN 1 13

DOMAIN 14 34

DOMAIN 35 54

DOMAIN 55 75

DOMAIN 76 90

DOMAIN 91 111

DOMAIN 112 188

DOMAIN 189 209

DOMAIN 210 221

CARBOHYD 139 139

SEQUENCE 221 AA; 24189 MW; E369888361AF4824 CRC64;

377 March 39.24; Score 462.5; DB 1; Length 221;

38 Local Similarity 41.6%; Pred. No. 1.7e-32;

39 Cons 94; Conservative 32; Mismatches 77; Indels 23; Gaps 5;

1 KYLLFVNFVFMALAGVITLVAALMLRDPOTTSLLYLKLNKPAKPTFYGGIYLLAVGA 60

12 KYLLFVNFVFMALAGVITLVAALMLRDPOTTSLLYLKLNKPAKPTFYGGIYLLAVGA 67

61 VMHVFGLGCVGAIOESQCLGTFEFLGVILFACVAVAGIAGFVFNKQIAKDVYQFIDOA 120

68 LMAVAVGPGCGMREOCVAGSFFTLVIFAAEVTVGPAFGVAVARHVQSMTEEA 127

121 LQAQVMDDDANNAKAVVKTETHTLNCSSNALTTTLTTLRNLSLCPSS--GNNILFPLQ 177

128 YSDYVDRGRGNGTLI--TFHSAVQCCGKSSSEGVQPT-----CPRELPG-----H 171

178 QDCHQKIDELPSGLYIIGIAIAVAVIMFEMILSNVTLCCGIRNS 223

172 KNCIDIKETIISVLAOLIGIVGIGIAGLTIFGMIFFSWVLCAIRNS 217

MOUSE

TSN2 MOUSE STANDARD; PRT; 221 AA.

Q92206; Q92397;

28-FEB-2003 (Rel. 41, Created)

28-FEB-2003 (Rel. 41, Last sequence update)

15-SEP-2003 (Rel. 42, Last annotation update)

Tetraspanin 2 (Tspan-2).

TSPAN2.

Mus musculus (Mouse).

OC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OC NCBI\_TaxID=10090;

OX

[11]

SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Medulla oblongata;

RC MEDLINE=C57BL/6J; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishi Y.,

RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamada I.,

RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,

RA Schiraldi L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Guetincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilming L.,

RA Wyshaw-Borls A., Yoshida K., Haegawa Y., Kawaji H., Kohlsanki S.,

RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:685-690(2001).

RN

SEQUENCE FROM N.A.

RP MEDLINE=22388257; PubMed=12477932;

RX Strausberg R.L., Pelinold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buettner K.H., Schaefer C.F., Heath N.K.,

RA Hopkins R.F., Jordan K.C., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Udell T.B., Toohyuk S., Carninci P., Prange C.,

RA Raba S.S., Loguanello N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Boak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length

human and mouse cDNA sequences.";

RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

-1- FUNCTION: May play a role in signalling in oligodendrocytes in the

early stages of their terminal differentiation into myelin-forming

glia and may also function in stabilizing the mature sheath (By

similarity).

-1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

-1- SIMILARITY: Belongs to the tetraspanin (TM4SF) family.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

EMBL: AK018178; BAB31109.1; -

DR EMBL: BC007185; AAH07185.1; -

DR MGD: MGI:1917997; 6330415F13R1X.

DR InterPro: IPR000301; Transmem\_4.

DR Pfam: PF00335; transmembrane4; 1.

DR PRINTS: PR00259; TMFOUR.

DR PROSITE: PS00421; TM4.1; 1.

DR Glycoprotein; Transmembrane.

KM



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

Protein - protein search, using sw model -

Date: December 9, 2003, 09:35:04 / Search time 85 Seconds

(Without alignments)  
686.116 Million cell updates/sec

US-10-035-914-2

Sequence: 1 KTLVFPNFWFLAGGVILG.....IFEMILSMVLCCGIRNSVY 226

Gapop 10.0, Gapext 0.5

Number of hits satisfying chosen parameters: 830525

Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SPREMBL\_23.\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_prodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeal:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Query	Match	Length	DB	ID	Description
1	1180	100.0	236	11	Q91V78 mus musculus
2	1018	86.3	211	6	Q96L69
3	738	62.5	236	13	Q96L69
4	616.5	52.2	228	13	Q96L69
5	523.5	44.4	224	13	Q96L69
6	519.5	44.0	226	6	Q96L69
7	486	41.2	222	13	Q96L69
8	422.5	35.8	199	11	Q96L69
9	411.5	34.9	217	11	Q96L69
10	300	25.4	235	11	Q96L69
11	297	25.2	235	11	Q96L69
12	290.5	24.6	268	5	Q96L69
13	288	24.4	248	4	Q96L69
14	283.5	24.0	247	13	Q96L69
15	274.5	23.3	239	11	Q96L69
16	272	23.1	236	4	Q96L69

Query	Match	Length	DB	ID	Description
17	269.5	22.8	230	13	Q96L69
18	267.5	22.7	267	5	Q96L69
19	260.5	22.1	253	11	Q96L69
20	259	21.9	267	4	Q96L69
21	257	21.8	240	11	Q96L69
22	255.5	21.7	282	5	Q96L69
23	251.5	21.3	253	11	Q96L69
24	242	20.5	283	11	Q96L69
25	239.5	20.3	230	13	Q96L69
26	234	19.8	270	4	Q96L69
27	231	19.6	239	5	Q96L69
28	231	19.6	270	4	Q96L69
29	230	19.5	270	11	Q96L69
30	227.5	19.3	233	5	Q96L69
31	220	18.6	270	11	Q96L69
32	220	18.6	270	11	Q96L69
33	219.5	18.6	236	5	Q96L69
34	218.5	18.5	223	5	Q96L69
35	211.5	17.9	227	5	Q96L69
36	211	17.9	232	5	Q96L69
37	210.5	17.8	232	5	Q96L69
38	207	17.5	222	5	Q96L69
39	207	17.5	249	11	Q96L69
40	206	17.5	266	11	Q96L69
41	204	17.3	239	5	Q96L69
42	202	17.1	263	4	Q96L69
43	201.5	17.1	276	5	Q96L69
44	199.5	16.9	244	5	Q96L69
45	199	16.9	291	5	Q96L69

## ALIGNMENTS

## RESULT 1

Q91V78 PRELIMINARY; PRT; 236 AA.  
ID Q91V78  
AC Q91V78  
DT 01-DEC-2001 (TREMELREL. 19, Created)  
DT 01-DEC-2001 (TREMELREL. 19, Last sequence update)  
DT 01-MAR-2003 (TREMELREL. 23, Last annotation update)  
DE Similar to CD81 antigen (Target of antiproliferative antibody 1)  
DE (Tapa-1 protein)  
GN TREM5 OR CD81 OR TAPA-1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10990;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast tumor;  
RA Strauberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129 SV;  
RX MEDLINE=20377495; PubMed=10915772;  
RA Paulsen M., El-Maari O., Engemann S., Stroedicke M., Franck O.,  
RA Davies K., Reinhardt R., Reik W., Walter J.;  
RT "Sequence conservation and variability of imprinting in the Beckwith-  
Wiedemann syndrome gene cluster in human and mouse.";  
RL Hum. Mol. Genet. 9:1829-1841(2000).  
DR EMBL; BC011433; AA011433.1; -;  
DR EMBL; AJ251835; CAB94774.1; -;  
DR MGI; MGI:1096398; Cd81.  
DR MGI; MGI:1861718; Trpm5.  
DR InterPro: IPR000301; Transmem\_4.  
DR PRINTS; PR00259; TMFOUR.  
DR PROSITE; PS00421; TM4\_1; 1.  
SQ SEQUENCE 236 AA; 25814 MW; 7BDCB792E7F9FAE CRC64;  
Query Match 100.0%; Score 1180; DB 11; Length 236;  
Best Local Similarity 100.0%; Pred. No. 5.2e-105;

```

ches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 KYLLFVNFVFMLAGCVIIGVALMLRHDPTTSLYLKGNKPAKPTFYVGIYIILAVGA 60
|||||
11 KYLLFVNFVFMLAGCVIIGVALMLRHDPTTSLYLKGNKPAKPTFYVGIYIILAVGA 70
|||||
61 VMVFGVGCYGAIOESQCLGTFPTCLVILFACVAAAGWGFVNKQIAKDKQFYDQA 120
|||||
71 VMVFGVGCYGAIOESQCLGTFPTCLVILFACVAAAGWGFVNKQIAKDKQFYDQA 130
|||||
121 LQQAAMDDANNAKAVKTFHETLNCSSNALTTTLTILRNSLCPSGNITLPLQQDC 180
|||||
131 LQQAAMDDANNAKAVKTFHETLNCSSNALTTTLTILRNSLCPSGNITLPLQQDC 190
|||||
181 HKIKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 226
|||||
191 HKIKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 236
|||||

T 2
9
09GLG9 PRELIMINARY; PRT; 211 AA.
09GLG9;
01-MAR-2001 (TrEMBLrel. 16, Created)
01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
CD81 (Fragment).
Saguinus oedipus (Cotton-top tamarin).
Bukarya; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Saguinus.
NCBI_TaxID=9490;
[1]
SEQUENCE FROM N.A.
MEDLINE=20534982; PubMed=11080483;
Allander T., Fornis X., Emerson S.U., Purcell R.H., Bukh J.;
"Hepatitis C virus envelope protein E2 binds to CD81 of tamarins.";
Virology 277:358-367(2000)
EMBL; AF274885; AAC31390.1; -
InterPro; IPR000301; Transmem_4.
Pfam; PF00335; transmembrane4; 1.
PRINTS; PR00259; TMFOUR.
PROSITE; PS00421; TM4_1; 1.
NON_TER 1 211
SEQUENCE 211 AA; 23091 MW; CACB34E89DAD3BD CRC64;

CY Match 86.3%; Score 1018; DB 6; Length 211;
c Local Similarity 91.0%; Pred. No. 1.5e-89;
ches 192; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

4 LKVFNFVFMLAGCVIIGVALMLRHDPTTSLYLKGNKPAKPTFYVGIYIILAVGA 63
|||||
1 LKVFNFVFMLAGCVIIGVALMLRHDPTTSLYLKGNKPAKPTFYVGIYIILAVGA 60
|||||
64 PVGFGVGCYGAIOESQCLGTFPTCLVILFACVAAAGWGFVNKQIAKDKQFYDQA 123
|||||
61 PVGFGVGCYGAIOESQCLGTFPTCLVILFACVAAAGWGFVNKQIAKDKQFYDQA 120
|||||
124 AVMDDDANNAKAVKTFHETLNCSSNALTTTLTILRNSLCPSGNITLPLQQDC 183
|||||
121 AVMDDDANNAKAVKTFHETLNCSSNALTTTLTILRNSLCPSGNITLPLQQDC 180
|||||
184 IDELPSGKLYLIGIAIVAVIMIFEMILSM 214
|||||
181 IDELPSGKLYLIGIAIVAVIMIFEMILSM 211
|||||

T 3
7
39DSD7 PRELIMINARY; PRT; 236 AA.
39DSD7;
01-MAR-2001 (TrEMBLrel. 16, Created)
01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

```

```

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CD81.
GN CD81.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
NCBI_TaxID=7955;
RN [1]
RP
SEQUENCE FROM N.A.
RX MEDLINE=21100417; PubMed=11167010;
RA Yoder J.A., Litman G.W.;
RT "The zebrafish fltn1, slc3a2, men1, pc, fgf3 and cycd1 genes define two
RT regions of conserved synteny between linkage group 7 and human
RT chromosome 11q13."
RL Gene 261:235-242(2000).
DR EMBL; AF295377; AAC37840.1; -
DR ZFIN; ZDB-GENE-000831-5; cd81.
DR InterPro; IPR01092; HLH_Basic.
DR Pfam; PF00335; transmembrane4; 1.
DR PRINTS; PR00259; TMFOUR.
DR PROSITE; PS00038; HLH_1; 1.
DR PROSITE; PS00421; TM4_1; 1.
SQ SEQUENCE 236 AA; 26168 MW; 3C9964B8FD953428 CRC64;

Query Match 62.5%; Score 738; DB 13; Length 236;
Best Local Similarity 64.0%; Pred. No. 1.1e-62;
Matches 146; Conservative 24; Mismatches 52; Indels 6; Gaps 3;

QY 1 KYLLFVNFVFMLAGCVIIGVALMLRHDPTTSLYLKGNKPAKPTFYVGIYIILAVGA 60
|||||
13 KYLLFVNFVFMLAGCVIIGVALMLRHDPTTSLYLKGNKPAKPTFYVGIYIILAVGA 72
|||||
61 VMVFGVGCYGAIOESQCLGTFPTCLVILFACVAAAGWGFVNKQIAKDKQFYDQA 120
|||||
73 VMVFGVGCYGAIOESQCLGTFPTCLVILFACVAAAGWGFVNKQIAKDKQFYDQA 132
|||||
121 LQQAAM--DDANNAKAVKTFHETLNCSSNALTTTLTILRNSLCPSGNITLPLQQ 178
|||||
133 YKGAATYVNDNNPNTAVLAKVHETLQCGKNLTALVDRMLTITCPB--HLRTAV-- 188
|||||
189 DCHTEIKMLFTDKISLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 236
|||||

QY 179 DCHOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 226
|||||
189 DCHTEIKMLFTDKISLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 236
|||||

DB 189 DCHTEIKMLFTDKISLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 236
|||||

QY 179 DCHOKIDELFSGKLYLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 226
|||||
189 DCHTEIKMLFTDKISLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 236
|||||

DB 189 DCHTEIKMLFTDKISLIGIAIVAVIMIFEMILSMVLCGGIRNSSVY 236
|||||

RESULT 4
Q8AV92 PRELIMINARY; PRT; 228 AA.
AC Q8AV92;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE CD9-like protein.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
NCBI_TaxID=757;
RN [1]
RP
SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=22295056; PubMed=12391333;
RA Ulnak-Ool T., Mayer W.B., Sato A., Dongak R., Cooper M.D., Klein J.;
RT "Lamprey lymphocyte-like cells express homologs of genes involved in
RT immunologically relevant activities of mammalian lymphocytes.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14356-14361(2002).
DR EMBL; AY152677; AAN64299.1; -
SQ SEQUENCE 228 AA; 25331 MW; DDE0EA31B988935 CRC64;

```

```

Query Match 52.2%; Score 616.5; DB 13; Length 228;
Best Local Similarity 51.3%; Pred. No. 4.7e-51;
Matches 116; Conservative 45; Mismatches 56; Indels 9; Gaps 5;

```



1 KILLFVNFVFWLGGVILGVALMLRHPDPTSLYLTLKPKPAFTFYGIYILAVGA 60  
 11 KILLFVNFVFWLGGVILGVALMLRHPDPTSLYLTLKPKPAFTFYGIYILAVGA 66  
 61 VMMFVGLGCGYAGIIOESQCLGTFCTCLVILFACVAGIWFVFNKQIADVKQFYDQA 120  
 67 LMMVGLGCGYAGIIOESQCLGTFCTCLVILFACVAGIWFVFNKQIADVKQFYDQA 125  
 121 LQOAVMDDDANNAKAVKTFHETLNCSSNALTTTLTLNLSLCPSSGNILTPILQDC 180  
 126 LVHEHKKTDSSKALITFHTLSCCGMVTITNLDSDFKD-ICPT-OKTIDPTV--DC 181  
 181 HOKIDELFSGKLYIGIAIIVAVIMIFEMILSMVLCGIRNSVY 226  
 182 IROIDEIFTSKYIACIGIGIVIMIFGMIFSMALCCAVRNRRDY 227

LT 5

IC9 PRELIMINARY; PRT; 224 AA.

091BC9  
 091BC9; 01-OCT-2000 (Tremblrel. 15, Created)  
 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 CD9 antigen.

CD9.  
 Gallus gallus (Chicken).

Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 Gallus.

NCBI\_TaxID=9031;

[1]

SEQUENCE FROM N.A.

TISSUE=spinal cord;

Kobayashi T., Mekada E.;

"Molecular cloning and characterization of Chick CD9,"  
 Submitted (SFP-1999) to the EMBL/Genbank/DBJ databases.

EMBL; AB032767; BAA3011.1; -  
 InterPro; IPR000301; Transmem\_4.  
 Pfam; PR00335; transmembrane4; 1.  
 PRINTS; PR00259; TMFOUR.  
 PROSITE; PS00421; TM4\_1; 1.

SEQUENCE 224 AA; 24969 MW; AB64BDF849516151 CRC64;

Query Match 44.4%; Score 523.5; DB 13; Length 224;  
 Best Local Similarity 47.3%; Pred. No. 3.7e-42;

Matches 107; Conservative 31; Mismatches 69; Indels 19; Gaps 5;

1 KILLFVNFVFWLGGVILGVALMLRHPDPTSLYLTLKPKPAFTFYGIYILAVGA 60  
 11 KILLFVNFVFWLGGVILGVALMLRHPDPTSLYLTLKPKPAFTFYGIYILAVGA 66  
 61 VMMFVGLGCGYAGIIOESQCLGTFCTCLVILFACVAGIWFVFNKQIADVKQFYDQA 120  
 67 LMMVGLGCGYAGIIOESQCLGTFCTCLVILFACVAGIWFVFNKQIADVKQFYDQA 126  
 121 LQOAVMDDDANNAKAVKTFHETLNCSSNALTTTLTLNLSLCPSSGNILTPILQDC 177  
 127 YEKK----SOPAAEILKAFHFAINCCG-----ITGGLEQOIMESCPK-KTLPSFTT 174  
 178 QDCHQKIDELFSGKLYIGIAIIVAVIMIFEMILSMVLCGIRNS 223  
 175 MPEPKAIDVFNSTLAVGAGIIVIMIFGMIFSMALCCAVRNRRDY 220

DE Transmembrane protein CD9.  
 CN CD9.  
 OS Sus scrofa (pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Heinz M., Huang C.A., Emery D.W., Giovino M.A., LeGuern A.,  
 RA Kurilla-Mahon B., Theodore P., Arn J.S., Sykes M., Mulligan R.C.,  
 RA Down J.D., Sachs D.H., Goodell M.A.;  
 RT "Use of CD9 Expression to Enrich for Porcine Hematopoietic  
 RT Progenitors."  
 RT Exp Hematol. 0:0-0(2002).  
 DR EMBL; AF525029; AAB81376.1; -  
 DR InterPro; IPR000301; Transmem\_4.  
 DR PRINTS; PR00259; TMFOUR.  
 DR PROSITE; PS00421; TM4\_1; 1.  
 SQ SEQUENCE 226 AA; 25127 MW; 885C244477F95C23 CRC64;

Query Match 44.0%; Score 519.5; DB 6; Length 226;  
 Best Local Similarity 46.6%; Pred. No. 9.1e-42;  
 Matches 104; Conservative 37; Mismatches 71; Indels 11; Gaps 5;

QY 1 KILLFVNFVFWLGGVILGVALMLRHPDPTSLYLTLKPKPAFTFYGIYILAVGA 60  
 DB 11 KILLFVNFVFWLGGVILGVALMLRHPDPTSLYLTLKPKPAFTFYGIYILAVGA 66  
 QY 61 VMMFVGLGCGYAGIIOESQCLGTFCTCLVILFACVAGIWFVFNKQIADVKQFYDQA 120  
 DB 67 LMMVGLGCGYAGIIOESQCLGTFCTCLVILFACVAGIWFVFNKQIADVKQFYDQA 126  
 QY 121 LQOAVMDDDANNAKAVKTFHETLNCSSNALTTTLTLNLSLCPSSGNILTPILQDC 180  
 DB 127 YNKKKGDDPD--RETLLKALHYALDDCG---LMGVSQQLAD-ICPD-KDVLSSLPKPC 179  
 QY 181 HOKIDELFSGKLYIGIAIIVAVIMIFEMILSMVLCGIRNS 223  
 DB 180 PEALKEVFNKFNHILGAVGIGIIVAVIMIFGMIFSMALCCAVRNRRDY 222

RESULT 7

QY 08AV91 PRELIMINARY; PRT; 222 AA.

AC 08AV91; 01-MAR-2003 (Tremblrel. 23, Created).

DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)

DB CD9-like protein.

OS Petromyzon marinus (Sea lamprey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;

OC Petromyzontiformes; Petromyzontidae; Petromyzon.

OX NCBI\_TaxID=7757;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=lymphocyte;

RX MEDLINE=2295056; PubMed=12391333;

RA Unuk-Ooi T., Mayer W.E., Sato A., Dongak R., Cooper M.D., Klein J.;

RT "Lamprey lymphocyte-like cells express homologs of genes involved in

immunologically relevant activities of mammalian lymphocytes."

Proc. Natl. Acad. Sci. U.S.A. 99:14356-14361(2002).

DR EMBL; AY152678; AAB64300.1; -  
 SQ SEQUENCE 222 AA; 24637 MW; CACOA561301EFB54 CRC64;

Query Match 41.2%; Score 486; DB 13; Length 222;  
 Best Local Similarity 43.5%; Pred. No. 1.4e-38;

Matches 100; Conservative 40; Mismatches 68; Indels 22; Gaps 6;

QY 1 KILLFVNFVFWLGGVILGVALMLRHPDPTSLYLTLKPKPAFTFYGIYILAVGA 60  
 DB 11 KILLFVNFVFWLGGVILGVALMLRHPDPTSLYLTLKPKPAFTFYGIYILAVGA 66  
 QY 61 VMMFVGLGCGYAGIIOESQCLGTFCTCLVILFACVAGIWFVFNKQIADVKQFYDQA 120

```

67 1MMVGGFLGCCGAIQBSQCMGTEFTLIVILFSDERBAVAFHDEVEFMLEYN-- 124
121 LQAVMDDDNNK-AVWT---FHEITNCCGSNAITLITLIRNSLCGSGNITPL 176
125 ---GLINDNSNSTRSASTKSLIHSITLDCGQPNLN--YTSILSTYSCP-----PEY 172
177 QODCHOKIDELFSGLKYLIGIAIVAVIMIFEMILSMVCCGIRNSSY 226
173 TRKCVDEIRSFGLTKYIYAGLAVAVIMIFGMIFMAVCAIRNDSPY 222

9
T 8
PRELIMINARY; PRT; 199 AA.
28C0C9;
01-MAR-2003 (Tremblrel. 23, Created)
01-MAR-2003 (Tremblrel. 23, Last sequence update)
01-MAR-2003 (Tremblrel. 23, Last annotation update)
TSPAN-2 protein homolog (Fragment).
Mus musculus (Mouse)
Bakayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=Testis;
MEDLINE=22354683; PubMed=12466851;
The RIKEN Genome Exploration Research Group Phase I & II Team;
"Analysis of the mouse transcriptome based on functional annotation of
50,770 full-length cDNAs.";
Nature 420:563-573(2002).
EMBL; AK031683; BAC27511.1; -.
NON_TER
3SEQUENCE 199 AA; 2161 MW; 5DD1929442680978 CRC64;

CY Match 35.8%; Score 422.5; DB 11; Length 199;
Local Similarity 40.0%; Pred. No. 1.6e-32;
ches 86; Conservative 31; Mismatches 75; Indels 23; Gaps 5;

12 WLAGVILGVALMLRHPDPTSLIYELGKRPAPNTEFYGYILIVAGVMMVFGCG 71
1 WLAGSAVIAFGMLFRGGTMKDL-----SSEDKSPRYFYGLVYLVGAGALMTVGFCCG 56
72 GAIOBQCLLGTFTCLVILFACVAAAGTGWGPNKQDIADVQFTDQALQAVMDDDAN 131
57 GMRBESQCVLGSFPTCLVIFAAVLTGVFAFGKQVAIRHVSMTYEAASDYLDKDRAG 116
132 NAAVAVKTHETLNCGGSNAITLITLIRNSLCPS---GAILITPLQODCHOKIDELF 188
117 NGTLI--THSAFQCCGKSSBOVPT-----CPKELPG-----HNKCIDKLETYI 160

189 SGKLYLIGIAIVAVIMIFEMILSMVCCGIRNS 223
161 SAKQLIGIVIGIAGLTIFGMIFSMVLCALIRNS 195

3
f 9
PRELIMINARY; PRT; 217 AA.
29D1X8;
01-JUN-2001 (Tremblrel. 17, Created)
01-JUN-2001 (Tremblrel. 17, Last sequence update)
01-OCT-2002 (Tremblrel. 22, Last annotation update)
3230119D02RIK protein.
3230119D02RIK.
Mus musculus (Mouse)
Bakayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.

```

```

RC STRAIN=C57BL/6J; TISSUE=Brain;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai T., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Akawa T., Hara A., Fukumishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H. A., Asbunner M., Batalov S., Casavant T.,
RA Pletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Koenig P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L. M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M. F.,
RA Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinich S., Hill D., Hofmann M., Hume D. A., Kamita M., Lee N. H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombere P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schenbach C., Seta T., Shibata Y., Storch K. F.,
RA Suzuki H., Toyooka K., Wang K. H., Weltz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Haegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK020982; BAB32269.1; -.
DR MGI; MGI:1925095; B230119D02RIK.
DR InterPro; IPR000301; Transmem. 4.
DR Pfam; PF00335; Transmembrane4; 1.
DR PRINTS; PR00259; TMFOUR.
DR PROSITE; PS00421; TM4_1; 1.
SQ SEQUENCE 217 AA; 23448 MW; E18DB13DD73055CE CRC64;

Query Match 34.9%; Score 411.5; DB 11; Length 217;
Best Local Similarity 39.7%; Pred. No. 1.9e-31;
Matches 85; Conservative 31; Mismatches 75; Indels 23; Gaps 5;

QY 13 LAGVILGVALMLRHPDPTSLIYELGKRPAPNTEFYGYILIVAGVMMVFGCG 72
DB 20 LAGSAVIAFGMLFRGGTMKDL-----SSEDKSPRYFYGLVYLVGAGALMTVGFCCG 75
QY 73 AIOBQCLLGTFTCLVILFACVAAAGTGWGPNKQDIADVQFTDQALQAVMDDDAN 132
DB 76 AMRBSQCVLGSFPTCLVIFAAVLTGVFAFGKQVAIRHVSMTYEAASDYLDKDRAGN 135
QY 133 AAVVAVKTHETLNCGGSNAITLITLIRNSLCPS---GAILITPLQODCHOKIDELF 189
DB 136 GTLI--THSAFQCCGKSSBOVPT-----CPKELPG-----HNKCIDKLETYI 179
QY 190 GLVYLIGIAIVAVIMIFEMILSMVCCGIRNS 223
DB 180 AKQLIGIVIGIAGLTIFGMIFSMVLCALIRNS 213

RESULT 10
ID 055158 PRELIMINARY; PRT; 235 AA.
AC 055158;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE D6.1A protein.
OS Rattus norvegicus (Rat).
OC Bakayota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BD IX;
RX MEDLINE=98198483; PubMed=9531564;
RA Claas C., Selter S., Claas A., Savelyeva L., Schwab M., Zoeller M.;
RT "Association between the Rat homologue of CO-029, a metacatalis-
associated tetraepanin molecular and consumption coagulopathy.";
J. Cell Biol. 141:267-280(1998).
[2]
RN
RP SEQUENCE FROM N.A.

```



Svirskas R., Tector C., Turner R., Venter B., Wang A.H., Wang X., Wang Z.-Y., Maasman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodard T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.W., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers B.W., Rubin G.M., Venter J.C. /  
 "The genome sequence of *Drosophila melanogaster*." /  
 Science 287:2185-2195(2000).

## SEQUENCE FROM N.A.

Tsang G., Brokstein P., Friese E., Harvey D., Evans-Holm M., Lewis S.B., Sub C., Rubin G.M. /  
 "Full length *Drosophila melanogaster* cDNA sequence." /  
 Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 EMBL: A5003753; AAF56520.1; -  
 EMBL: AF160941; AAD46881.1; -  
 Flybase: FBgn0027865; Tsp96F.  
 InterPro: IPR000301; Transmem. 4.  
 Pfam: PF00335; transmembrane4; 1.  
 PRINTS: PR00259; TMFOUR.  
 SEQUENCE 268 AA; 29849 MW; 49CPD43591A1B712 CRC64;

TY Match 24.6%; Score 290.5; DB 5; Length 268;  
 Local Similarity 25.9%; Pred. No. 9.5e-20;  
 Conservative 50; Mismatches 94; Indels 51; Gaps 7;

```

1 KYLLFVNFVFMAGVILGVALLMRHPDPTSLLYLELNKPAKPAFTVGYIYLAVGA 60
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
11 KYLMVFNILFMILGTLTVTSVMMLDP--TFMLSTWNTQ---NHVHALVFLAIGI 64
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61 VMVFPGLCYGALIOESQCLGTFPTCLVILFACVAVAGIWFVFNKDIADVKQFYDQ 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
65 LITLGAFFPCCGCRSSQCLIVSFFCVIILVMAQIAGAMAHAKDLDIVAAVKSS 124
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
121 LQQAAMDDANNAKAVKTFHETLNCGSNA--LTLTTLRN----- 162
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
125 VQERYQSTSSRTVTFDTLQKNLKCCGADGPDWATSRFNVNDRINVEIAVSNVVFY 184
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
163 ----SLCP-----GNIILFPLIQDCHQKIDELFSGKYL--LIGIAIIV 201
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
185 NIPESCCKNLKNCELSRLKFGFGLNNAIVQGCVDLBI---IYENVITFAVT 240
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
202 VAVIM--IFEMILSMVLCGIRN 222
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
241 AVVILIELSLTFLSLCCAVRN 263
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

3 PRELIMINARY; PRT; 248 AA.  
 396SJ8  
 396SJ8:

01-DEC-2001 (TREMBlrel. 19, Created)  
 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 Hypothetical protein FLJ14809.  
 Homo sapiens (Human).  
 Chkaryocyt; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;

## SEQUENCE FROM N.A.

Isogai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K., Tanai H., Kimata M., Watanabe M., Hirooka S., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuko Y., Kanehori K. /  
 "NEO human cDNA sequencing project." /  
 Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
 EMBL: AK027715; BAB55318.1; -  
 InterPro: IPR000264; Serum\_albumin.  
 InterPro: IPR000301; Transmem. 4.  
 Pfam: PF00335; transmembrane4; 1.  
 PRINTS: PR00259; TMFOUR.

DR PROSITE: PS00212; ALBUMIN; 1.  
 DR PROSITE: PS00421; TM4\_1; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 248 AA; 27710 MW; 6E2869E9D1D35C8E CRC64;

Query Match 24.4%; Score 288; DB 4; Length 248;  
 Best local similarity 30.0%; Pred. No. 1.5e-19;  
 Matches 75; Conservative 44; Mismatches 91; Indels 40; Gaps 7;

```

QY 1 KYLLFVNFVFMAGVILGVALLMRHPDPTSLLYLELNKPAKPAFTVGYIYLAVGA 60
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 10 KYLMVFNILFMILGTLTVTSVMMLDP--TFMLSTWNTQ---NHVHALVFLAIGI 63
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 VMVFPGLCYGALIOESQCLGTFPTCLVILFACVAVAGIWFVFNKDIADVKQFYDQ 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 64 LITLGAFFPCCGCRSSQCLIVSFFCVIILVMAQIAGAMAHAKDLDIVAAVKSS 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 LQQAAMDDANNAKAVKTFHETLNCGSNA-----LTLTTLRN----- 157
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 121 LKRYQGNNDIVFSAITWNSVMTITFGCCGVNGEPDFKASVRLTLTDEEVPACCRRE 180
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 158 -----TLRNSLCPSSGNIILFPLIQDCHQKIDELFSGKYLIGIAIIVAVIMIFEMI 211
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 181 PQSRDGVILSRRECLGRSLF--LAKQCYTVILMTFTVYVILAGALNIGVLAIEFLPMI 238
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 212 LSMVLCGIR 221
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 239 FAMCLFRGIQ 248
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

## RESULT 14

Q9PTB0 PRELIMINARY; PRT; 247 AA.

AC Q9PTB0;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE Neuronal tetraspanin.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 ON NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=spinal cord;  
 RX MEDLINE=20029748; PubMed=10561501;  
 RA Peron J., Bixby J.L.;  
 RT "Tetraspanins expressed in the embryonic nervous system." /  
 RL FEBS Lett. 461:86-90(1999).  
 DR EMBL: AF206661; AAF19031.1; -  
 DR InterPro: IPR00264; Serum\_albumin.  
 DR InterPro: IPR000301; Transmem. 4.  
 DR Pfam: PF00335; transmembrane4; 1.  
 DR PRINTS: PR00259; TMFOUR.  
 DR PROSITE: PS00212; ALBUMIN; 1.  
 DR PROSITE: PS00421; TM4\_1; 1.  
 SQ SEQUENCE 247 AA; 27804 MW; D571A0D617E11AF CRC64;

Query Match 24.0%; Score 283.5; DB 13; Length 247;

Best local similarity 30.1%; Pred. No. 4.1e-19;  
 Matches 75; Conservative 44; Mismatches 91; Indels 39; Gaps 7;

```

QY 1 KYLLFVNFVFMAGVILGVALLMRHPDPTSLLYLELNKPAKPAFTVGYIYLAVGA 60
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 10 KYLMVFNILFMILGTLTVTSVMMLDP--TFMLSTWNTQ---NHVHALVFLAIGI 63
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 61 VMVFPGLCYGALIOESQCLGTFPTCLVILFACVAVAGIWFVFNKDIADVKQFYDQ 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 64 LITLGAFFPCCGCRSSQCLIVSFFCVIILVMAQIAGAMAHAKDLDIVAAVKSS 120
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 121 LQQAAMDDANNAKAVKTFHETLNCGSNA-----LTLTTLRN----- 162
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

